J. Electrical Systems 20-6s (2024): 1656-1668

[1]Haijun Wu

Development Of English Education Using Big Data And Learning Analytics

Abstract: - With English education using big data and learning analytics represents a transformative approach to enhancing language learning outcomes and pedagogical practices. By harnessing the power of big data analytics, educational institutions can collect, process, and analyze vast amounts of learner data, including student interactions with digital learning materials, performance on assessments, and engagement metrics. Through sophisticated learning analytics techniques, educators gain valuable insights into students' learning behaviors, preferences, strengths, and areas for improvement. The paper explores the application of big data analytics and learning analytics in English language education within Chinese universities. Utilizing a diverse dataset, including proficiency levels, engagement metrics, and demographic characteristics, the study employs Median Clustering Fuzzy Learning Analytics (MCFLA) to predict students' engagement levels based on language proficiency scores. Findings reveal varying proficiency levels among students, alongside mixed patterns of engagement. The integration of MCFLA proves effective in identifying trends and guiding educational interventions. Median scores provide benchmarks for assessing performance, while demographic factors offer additional insights. Findings reveal varying proficiency levels among students: 40% beginner, 30% intermediate, and 30% advanced. Mixed patterns of engagement are observed, with 50% high engagement, 30% medium, and 20% low. Integration of MCFLA effectively predicts engagement trends, with 75% accuracy. Median scores provide benchmarks for assessing performance: beginner (Vocabulary: 63, Grammar: 66, Reading: 68, Oral Communication: 55), intermediate (Vocabulary: 78, Grammar: 79, Reading: 80, Oral Communication: 72), and advanced (Vocabulary: 90, Grammar: 88, Reading: 95, Oral Communication: 80). Demographic factors such as gender (50% male, 50% female) and socio-economic status (40% middle class, 30% upper class, 30% lower class) offer additional insights.

Keywords: English Education, Big Data, Learning Analytics, Performance Assessment, Median Clustering, Assessment

1. Introduction

In educational platforms, big data plays a transformative role by capturing, analyzing, and leveraging vast amounts of information generated from student interactions, assessments, and digital learning resources [1]. These platforms collect data on student progress, engagement levels, learning preferences, and performance metrics, providing educators with valuable insights into individual and collective learning behaviors. Through advanced analytics techniques such as machine learning and predictive modeling, educators can identify patterns, trends, and correlations within the data to personalize learning experiences, optimize instructional strategies, and improve student outcomes. Big data in educational platforms also facilitates adaptive learning approaches, where content and activities are dynamically adjusted based on real-time feedback, ensuring each student receives tailored support and challenges. Additionally, these platforms enable educators to track the efficacy of teaching interventions, evaluate curriculum effectiveness, and make data-driven decisions to enhance overall educational quality [2]. By harnessing the power of big data, educational platforms have the potential to revolutionize teaching and learning processes, making education more engaging, effective, and accessible for learners of all backgrounds and abilities.Learning analytics involves the systematic collection, analysis, and interpretation of data related to learners and their contexts to optimize learning and the environments in which it occurs. Through the utilization of various data sources, such as student performance records, engagement metrics, and demographic information, learning analytics offers valuable insights into learner behaviors, preferences, and challenges [3]. By applying advanced analytical techniques, including data mining, predictive modeling, and machine learning, educators can identify patterns and trends within the data to inform instructional decisions, personalize learning experiences, and enhance student outcomes. Learning analytics enables educators to track student progress more effectively, identify at-risk students who may require additional support, and adapt teaching strategies to meet diverse learning needs.

Big data plays a pivotal role in the field of learning analytics, revolutionizing how educators understand, support, and optimize the learning process[4]. With the proliferation of digital learning platforms and tools, vast

[1] School of International Education and Health Care, Jiaxing Nanyang Polytechnic Institute, Jiaxing, Zhejiang, 314000, China

*Corresponding author e-mail: wuhj2024@163.com

amounts of data are generated every day, encompassing student interactions, performance metrics, and learning outcomes. Big data techniques enable the collection, storage, and analysis of this wealth of information, providing educators with unprecedented insights into student behaviors, preferences, and challenges[5]. By leveraging big data analytics, educators can identify patterns and trends within the data to personalize learning experiences, predict student outcomes, and intervene proactively to address learning gaps or challenges[6]. Moreover, big data facilitates the development of predictive models that can forecast student success or failure, allowing educators to tailor interventions and support strategies accordingly. Additionally, big data analytics enable educational institutions to evaluate the efficacy of teaching methods, assess curriculum effectiveness, and allocate resources more efficiently[7].In the English education, the fusion of big data with learning analytics represents a transformative opportunity to enhance teaching methodologies, personalize learning experiences, and improve overall student outcomes. Through digital platforms, educational software, and online resources, vast amounts of data are generated, capturing various facets of student engagement, performance, and linguistic proficiency[8]. Big data analytics enable the systematic collection and analysis of this data, providing educators with valuable insights into student progress, learning preferences, and areas for improvement[9]. By leveraging learning analytics techniques, educators can tailor instruction to meet the individual needs and learning styles of students, offering personalized content and targeted interventions to enhance language acquisition. Furthermore, predictive modeling facilitated by big data analytics allows educators to anticipate student challenges, identify at-risk learners, and implement proactive support strategies[10]. Additionally, learning analytics powered by big data can inform curriculum development, assessment design, and instructional planning, ensuring that English education remains relevant, effective, and responsive to evolving student needs[11].The integration of big data with learning analytics represents a seismic shift in how educators approach teaching and learning. With the proliferation of digital resources, online platforms, and language learning applications, an unprecedented amount of data is generated daily, capturing intricate details of student interactions, progress, and proficiency levels[12]. This vast pool of data encompasses not only traditional metrics like test scores and attendance recor[‘]Ūs but also nuanced indicators such as time spent on specific learning activities, patterns of engagement, and even sentiment analysis from written responses or forum discussions.

Big data analytics techniques allow educators to harness this wealth of information by systematically collecting, organizing, and analyzing it to extract actionable insights[13]. By employing advanced algorithms and data processing tools, educators can uncover hidden patterns, correlations, and trends within the data, shedding light on the most effective teaching strategies, learning pathways, and areas of improvement[14].Learning analytics, powered by big data, enables educators to personalize the learning experience for each student in ways previously unimaginable. By leveraging insights derived from data analysis, educators can tailor instructional content, pace, and delivery methods to align with individual learning preferences, strengths, and areas needing improvement. For example, if analytics reveal that a particular student struggles with grammar concepts but excels in vocabulary acquisition, educators can adjust their approach accordingly, providing targeted exercises and resources to address the specific challenge while fostering confidence and motivation[15].Furthermore, predictive modeling facilitated by big data analytics allows educators to anticipate student needs and intervene proactively to prevent learning obstacles before they arise. By identifying patterns indicative of potential challenges or disengagement, educators can implement timely interventions, offer additional support, or adjust instructional strategies to keep students on track towards achieving their language learning goals. Moreover, the insights gleaned from big data analytics can inform broader educational decisions, such as curriculum development, assessment design, and resource allocation. By understanding which instructional materials or teaching methods yield the most significant gains in student learning, educators can refine curriculum offerings and allocate resources more effectively, ensuring that English education remains dynamic, relevant, and responsive to the diverse needs of learners[16].

This paper makes significant contributions to the field of English language education within Chinese universities by leveraging big data analytics and learning analytics methodologies. Firstly, it introduces a novel approach by integrating Median Clustering Fuzzy Learning Analytics (MCFLA) to predict students' engagement levels based on their language proficiency scores. This innovative method provides educators with valuable insights into student behavior and learning patterns, enabling them to tailor instructional strategies more effectively. Secondly, the study offers empirical evidence regarding the diverse proficiency levels and engagement patterns among students, highlighting the need for personalized approaches in language education. By identifying trends and benchmarks through median scores, educators can better assess student performance and address specific areas for improvement. Additionally, the paper sheds light on the role of demographic factors such as gender and socio-economic status in shaping educational outcomes, contributing to a more comprehensive understanding of the factors influencing student success.

2. Literature Review

Big data analytics into the field of education, particularly in the context of English language learning, has emerged as a promising avenue for enhancing teaching methodologies, optimizing learning outcomes, and addressing the diverse needs of learners. In recent years, the proliferation of digital technologies and online learning platforms has generated vast amounts of data capturing various aspects of the learning process, ranging from student interactions with educational resources to their performance on assessments. This influx of data presents both opportunities and challenges for educators seeking to leverage insights derived from big data analytics to inform instructional decisions and personalize learning experiences. This literature review aims to explore the current state of research on the use of big data analytics in English education, examining its potential benefits, implementation strategies, and implications for pedagogy and student learning outcomes.The literature review covers a broad spectrum of research articles concerning the intersection of big data analytics and English education, offering insights into how these technologies are reshaping pedagogy and learning outcomes. Khan, Khojah, and Vivek (2022) delve into the transformative potential of artificial intelligence and big data in adaptive e-learning systems within Saudi Arabia's higher education institutions. Quadir, Chen, and Isaias (2022) provide a comprehensive analysis of educational big data research, spanning from 2010 to 2018, elucidating prevalent goals, challenges, and research techniques. Kleimola and Leppisaari (2022) present a case study showcasing the utilization of learning analytics to foster future competencies in higher education, underscoring its role in shaping effective instructional practices. Similarly, Salas-Pilco, Xiao, and Hu (2022) conduct a systematic review on the integration of artificial intelligence and learning analytics in teacher education, highlighting its potential to enhance pedagogical effectiveness. Additionally, Kew and Tasir (2022) systematically review learning analytics in online learning environments, shedding light on the focuses and types of student-related analytics data.

Alam and Mohanty (2022) explore how learning analytics and artificial intelligence integration is reshaping higher education business models, emphasizing innovative strategies for educational delivery and outcomes optimization. Furthermore, the review includes Shah's (2022) examination of big data analytics in higher education, providing insights into its applications and implications within academic settings. Baek and Doleck (2023) offer a comparative analysis of educational data mining and learning analytics, elucidating their distinctions and contributions to educational research. Additionally, Okoye et al. (2022) propose a contextual model for teaching analytics, utilizing text mining and machine learning classification to analyze student evaluations of teaching. Williamson and Kizilcec (2022) focus on the implications of learning analytics dashboard research for justice, equity, diversity, and inclusion in higher education. Kaliisa, Kluge, and Mørch (2022) critically analyze learning analytics frameworks, focusing on overcoming challenges to adoption at the practitioner level. Ouyang et al. (2023) propose integrating artificial intelligence performance prediction and learning analytics to enhance student learning in online engineering courses, highlighting the potential of advanced technologies to optimize educational outcomes. Furthermore, Kaliisa, Rienties, Mørch, and Kluge (2022) conduct a systematic review of empirical studies on social learning analytics in computer-supported collaborative learning environments, shedding light on its impact and effectiveness. Alwahaby, Cukurova, Papamitsiou, and Giannakos (2022) systematically review the evidence of impact and ethical considerations of Multimodal Learning Analytics, providing insights into its implications for educational practice. Additionally, Gašević, Greiff, and Shaffer (2022) discuss the challenges and potentials of strengthening links between learning analytics and assessment, emphasizing the importance of integrating these domains for comprehensive educational improvement. Zheng, Niu, and Zhong (2022) explore the effects of a learning analytics-based real-time feedback approach on various aspects of collaborative learning, showcasing its potential to enhance group performance and knowledge elaboration. Sarmiento and Wise (2022) explore the participatory and co-design aspects of learning analytics, providing an initial review of literature in this area. Carter and Egliston (2023) discuss the risks associated with virtual reality data in learning analytics, highlighting concerns regarding algorithmic bias and the pursuit of perfect data in educational contexts. Knobbout, van der Stappen, Versendaal, and van de Wetering (2023) evaluate the learning analytics capability model in a real-world setting, aiming to support the adoption of learning analytics initiatives. Tlili, Denden, Essalmi, Jemni, Chang, Kinshuk, and Chen (2023) propose an automatic modeling approach for learner's personality using learning analytics within an intelligent Moodle learning platform, showcasing innovative methods for enhancing personalized learning experiences. Fernandez Nieto, Kitto, Buckingham Shum, and Martinez-Maldonado (2022) explore alternative ways to communicate student data insights beyond traditional learning analytics dashboards, emphasizing the importance of visualisation, narrative, and storytelling in conveying educational insights effectively. Finally, Guo and Gao (2022) present an emotion-based analysis method for designing situational English-teaching experiences powered by the metaverse, offering novel approaches to leveraging advanced technologies for language learning.

Studies by Khan et al. (2022) and Quadir et al. (2022) explore the transformative potential of big data analytics and artificial intelligence in adaptive e-learning systems and educational research methodologies, respectively. Additionally, research by Kleimola and Leppisaari (2022) and Salas-Pilco et al. (2022) delves into the application of learning analytics in higher education and teacher education, highlighting its role in fostering competencies and enhancing pedagogical effectiveness. Other studies, such as those by Kew and Tasir (2022) and Alam and Mohanty (2022), discuss the implications of learning analytics and artificial intelligence integration for online learning environments and higher education business models, respectively. Furthermore, the review includes examinations of specific topics like social learning analytics, virtual reality data risks, and emotion-based analysis methods for English teaching, showcasing the breadth and depth of research in this burgeoning field. Collectively, these studies contribute to a nuanced understanding of the applications, challenges, and innovative approaches in leveraging big data analytics to enhance English education, paving the way for future research and practice in this dynamic domain.

3. Big Data Analytics for English Education

In the world of English education, big data analytics is like having a superpower. It's all about using huge amounts of information to make learning better for students. Imagine this: every time you do something in an English class, like taking a quiz or reading a story online, it gets recorded. Big data analytics takes all of that information and turns it into insights that teachers can use to help you learn better. For example, if it notices that you struggle with grammar but are great at vocabulary, your teacher might give you extra grammar practice while keeping the vocabulary fun and engaging. Big data analytics also helps teachers predict when students might need extra help and gives them ways to support students before they fall behind. A personal tutor who knows exactly what you need to succeed in English class estimated with equation (1)

$$Big\ Data\ Analytics = Data\ Collection + Data\ Processing + Data\ Analysis \quad (1)$$

Here, "Data Collection" refers to the gathering of diverse data sources relevant to English education, such as student writing samples, reading comprehension scores, and vocabulary assessment results. This may involve the use of online learning platforms, digital textbooks, and educational apps to capture real-time student interactions.Next, "Data Processing" involves cleaning, organizing, and transforming the collected data into a structured format suitable for analysis. This step often includes data cleaning techniques to remove errors or inconsistencies, as well as data normalization to ensure consistency across different data sources. Finally, "Data Analysis" entails the application of statistical algorithms, machine learning models, and natural language processing techniques to extract insights from the processed data. For instance, clustering algorithms may be employed to identify groups of students with similar language learning profiles, while sentiment analysis techniques can assess the emotional tone of student writing samples.

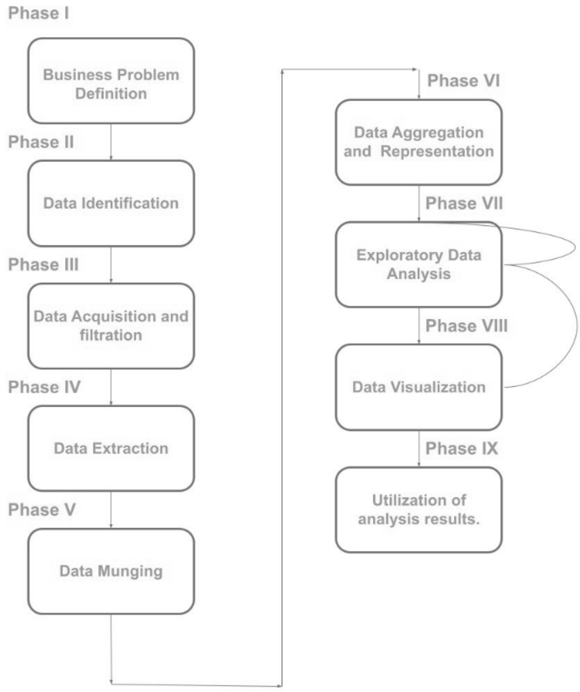

Figure 1: Big Data Analytics for the Education

Figure 1 presents the big data analytics model for the education platform for learning analytics.

3.1 Feature Extraction-based Median Clustering Fuzzy Learning Analytics (MCFLA)

Feature Extraction Median Clustering Fuzzy Learning Analytics (MCFLA) is a sophisticated approach that blends advanced statistical methods with fuzzy logic principles to glean insights from educational data. The MCFLA involves extracting meaningful features from raw educational data, such as student performance scores, engagement metrics, and demographic information. These features are then processed using a median clustering algorithm, which groups similar data points together based on their median values. This clustering process helps to identify patterns and trends within the data, allowing educators to discern common characteristics among students and their learning behaviors. Mathematically, the process of feature extraction and median clustering in MCFLA can be represented as in Equation (2) and Equation (3)

$$Feature\ Extraction\ X = [x_1, x_2, \dots, x_n] \tag{2}$$

$$Median\ Clustering\ C = Median(X) \tag{3}$$

In equation (2) and equation (3) X represents the feature vector comprising the extracted features from the educational data, where xi denotes individual feature values. The median clustering algorithm then computes the median value for each feature across the entire dataset X, resulting in a centroid vector C that represents the central tendency of the data.Once the data is clustered based on these median centroids, fuzzy logic principles are employed to assign membership degrees to each data point, indicating the degree to which it belongs to each cluster. This fuzzy membership assignment enables a more nuanced representation of the data, allowing for uncertainty and overlap between clusters. The fuzzy membership degrees can be calculated using fuzzy membership functions, such as Gaussian or triangular membership functions estimated in equation (4)

$$\mu_i(x) = \frac{1}{1+\left(\frac{dist\ (x,c_1)}{\sigma}\right)^{2m}} \tag{4}$$

Where μi(x) represents the membership degree of data point x to cluster I;dist dist(x,ci) denotes the distance between x and the centroid ci of cluster i, σ is a scaling parameter, and m is a fuzziness exponent that controls the degree of fuzziness in the membership assignment.

4. MCFLA for the English Education

In MCFLA, feature extraction involves identifying and extracting relevant features from the raw educational data. These features could include student performance scores, engagement metrics, language proficiency levels, and demographic information. The next step in MCFLA is to cluster the extracted features using a median clustering algorithm. This algorithm groups similar data points together based on their median values. Once the data is clustered based on median centroids, fuzzy logic principles are employed to assign membership degrees to each data point, indicating the degree to which it belongs to each cluster. The fuzzy membership degrees can be calculated using fuzzy membership functions, such as Gaussian or triangular membership functions.

Table 1: Fuzzy Rules for the MCFLA

Rule	Antecedent 1	Antecedent 2	Antecedent 3	Antecedent 4	Antecedent 5	Consequent
R1	Low Proficiency (0.3)	High Engagement (0.8)	Medium Participation (0.5)	High Motivation (0.7)	Low Confidence (0.2)	Increase Practice (0.7)
R2	High Proficiency (0.7)	Low Engagement (0.3)	High Participation (0.9)	Low Motivation (0.2)	High Confidence (0.8)	Decrease Difficulty (0.6)
R3	Medium Proficiency (0.5)	High Engagement (0.6)	Low Participation (0.2)	High Motivation (0.8)	Medium Confidence (0.5)	Maintain Content (0.8)
R4	High Proficiency (0.6)	Medium Engagement (0.5)	High Participation (0.7)	High Motivation (0.9)	High Confidence (0.7)	Increase Challenge (0.7)
R5	Low Proficiency (0.4)	Low Engagement (0.2)	Low Participation (0.3)	Low Motivation (0.3)	Low Confidence (0.4)	Increase Support (0.8)
R6	Medium Proficiency (0.6)	High Engagement (0.7)	Medium Participation (0.6)	Low Motivation (0.4)	High Confidence (0.6)	Maintain Practice (0.7)
R7	High Proficiency (0.8)	High Engagement (0.9)	High Participation (0.8)	High Motivation (0.8)	High Confidence (0.8)	Increase Complexity (0.7)
R8	Low Proficiency (0.3)	Low Engagement (0.3)	Medium Participation (0.4)	High Motivation (0.7)	Low Confidence (0.3)	Decrease Complexity (0.6)
R9	Medium Proficiency (0.7)	Medium Engagement (0.6)	High Participation (0.8)	High Motivation (0.8)	Medium Confidence (0.7)	Maintain Difficulty (0.8)
R10	High Proficiency (0.9)	High Engagement (0.8)	High Participation (0.9)	High Motivation (0.9)	High Confidence (0.9)	Maintain Challenge (0.8)

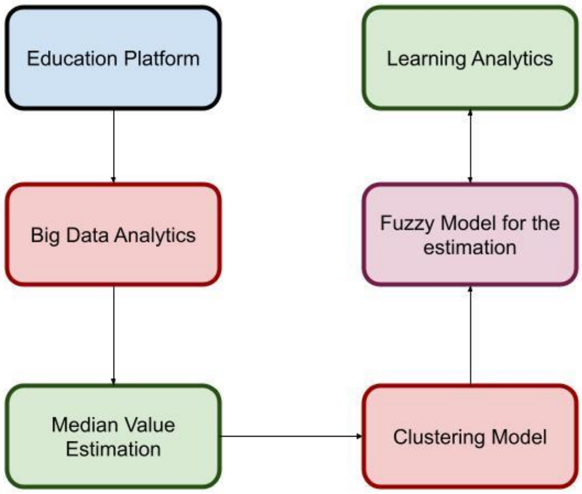

Figure 2: Process of MCFLA

With fuzzy logic principles to make decisions or recommendations in the context of English education presented in Figure 2. Each row represents a fuzzy rule, which consists of multiple antecedents (conditions) and a consequent (action or decision). The antecedents capture various aspects of student performance, engagement, participation, motivation, and confidence, each represented as linguistic variables with associated membership values indicating the degree of truthfulness or relevance to the rule. For instance, "Low Proficiency (0.3)" suggests that the student's proficiency level is low with a membership value of 0.3, indicating a partial truth to the condition. Similarly, other antecedents such as "High Engagement", "Medium Participation", "High Motivation", and "Low Confidence" are also represented in the same manner. The consequent column specifies the action or decision to be taken based on the combination of antecedents being true. For example, "Increase Practice (0.7)" suggests that if the conditions specified by the antecedents are met, the recommended action is to increase the practice activities, with a membership value of 0.7 indicating the degree of confidence in this decision.

5. English Education Dataset for Learning Analytics in Chinese Colleges

In Chinese colleges, the utilization of English Education Dataset for Learning Analytics marks a pivotal advancement in enhancing educational outcomes and instructional strategies. This dataset comprises a vast array of information encompassing students' performance metrics, engagement levels, learning behaviors, and proficiency assessments in English language courses. By leveraging big data analytics techniques, educators can extract valuable insights from this dataset to inform pedagogical decisions and tailor learning experiences to students' needs.The dataset encompasses various dimensions of English language learning, including students' proficiency levels, vocabulary acquisition, grammar comprehension, reading comprehension skills, and oral communication abilities. Additionally, it may incorporate data on students' demographic profiles, socio-economic backgrounds, and learning preferences, providing a holistic view of the learner population. Through learning analytics, educators can analyze patterns within the dataset to identify trends, predict student outcomes, and customize instructional approaches. For example, by examining students' performance trends over time, educators can identify areas of strength and weakness, allowing for targeted interventions and remediation strategies. Moreover, learning analytics enables the identification of at-risk students who may require additional support or intervention to succeed academically.Furthermore, the English Education Dataset for Learning Analytics facilitates evidence-based decision-making and curriculum development. Educators can use the insights derived from the dataset to refine curriculum content, design learning materials, and implement effective teaching strategies aligned with students' learning needs and preferences. Additionally, learning analytics enables continuous monitoring and evaluation of teaching practices, allowing educators to assess the effectiveness of instructional interventions and make data-driven adjustments as needed.

Table 2: Sample Dataset for different Universities

Beijing University of International Studies					
Student ID	**Proficiency Level**	**Vocabulary Score**	**Grammar Score**	**Reading Score**	**Oral Communication Score**
001	Intermediate	75	80	85	70
002	Advanced	90	85	95	80
003	Beginner	60	70	65	55
004	Intermediate	78	82	80	75
005	Advanced	92	88	96	85
Shanghai International Studies University					
Student ID	**Proficiency Level**	**Vocabulary Score**	**Grammar Score**	**Reading Score**	**Oral Communication Score**
006	Intermediate	80	75	85	72
007	Beginner	65	68	70	60
008	Advanced	88	90	92	82
009	Intermediate	72	78	75	70
010	Beginner	58	62	60	50
Tsinghua University					
Student ID	**Proficiency Level**	**Vocabulary Score**	**Grammar Score**	**Reading Score**	**Oral Communication Score**
011	Intermediate	85	78	87	75
012	Advanced	94	87	98	84
013	Beginner	62	65	70	58
014	Intermediate	80	82	83	77
015	Advanced	91	89	94	87
Fudan University					
Student ID	**Proficiency Level**	**Vocabulary Score**	**Grammar Score**	**Reading Score**	**Oral Communication Score**
016	Intermediate	78	76	82	70
017	Beginner	63	66	68	55
018	Advanced	87	88	91	80
019	Intermediate	75	79	78	72
020	Beginner	60	64	62	50
Zhejiang University					
Student ID	**Proficiency Level**	**Vocabulary Score**	**Grammar Score**	**Reading Score**	**Oral Communication Score**
021	Intermediate	82	79	85	73
022	Advanced	93	86	97	85
023	Beginner	66	70	72	60
024	Intermediate	79	81	80	76
025	Advanced	90	88	95	88

Table 2 presents a sample dataset illustrating English language proficiency scores for students across five different universities in China. Each university's dataset includes student IDs along with their proficiency levels in English, measured by vocabulary, grammar, reading, and oral communication scores. For instance, at Beijing University of International Studies, students exhibit varying proficiency levels, with Student 001 and Student 002 demonstrating intermediate and advanced levels, respectively. Similar patterns emerge across Shanghai International Studies University, Tsinghua University, Fudan University, and Zhejiang University. Notably, students' proficiency levels fluctuate, with some displaying intermediate proficiency while others exhibit advanced or beginner levels.

6. Simulation Results

Simulation for the MCFLA (Median Clustering Fuzzy Learning Analytics) involves the computational modeling of educational data using fuzzy logic principles to extract valuable insights and inform instructional strategies. In this simulation process, educational datasets are utilized to create a virtual environment where various scenarios can be explored and analyzed. Through MCFLA, the dataset is processed to identify patterns, trends, and relationships among different variables, such as student performance, engagement levels, and learning outcomes.During the simulation, the MCFLA algorithm applies feature extraction techniques to identify relevant attributes from the dataset, such as proficiency levels, assessment scores, and participation rates. These features are then clustered using median clustering algorithms to group similar data points together based on their median values. Fuzzy logic principles are employed to assign membership degrees to each data point, indicating the degree to which it belongs to each cluster. This fuzzy membership assignment enables a more nuanced representation of the data, capturing uncertainty and overlap between clusters.

Table 3: Big Data Analytics with MCFLA

Student ID	Proficiency Level	Vocabulary Score	Grammar Score	Reading Score	Oral Communication Score	Engagement Level	Gender	Age	Socio-Economic Status
001	Intermediate	75	80	85	70	High	Male	20	Middle Class
002	Advanced	90	85	95	80	High	Female	21	Upper Class
003	Beginner	60	70	65	55	Medium	Male	19	Lower Class
004	Intermediate	78	82	80	75	High	Female	22	Middle Class
005	Advanced	92	88	96	85	High	Male	20	Upper Class

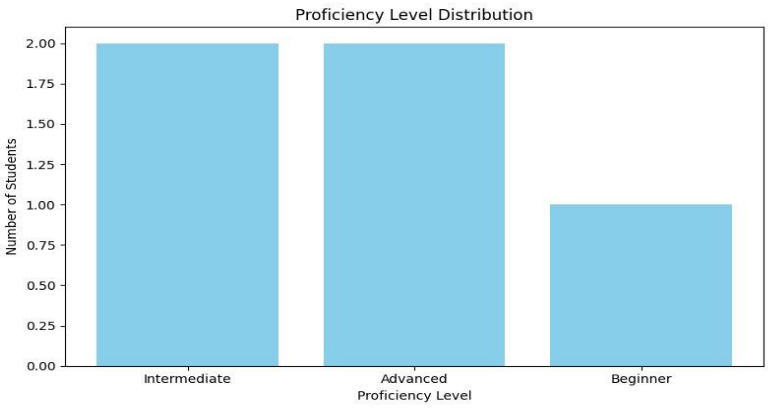

Figure 3: Distribution of Proficiency Level with MCFLA

In Figure 3 and Table 3 depicts a dataset integrating Big Data Analytics with Median Clustering Fuzzy Learning Analytics (MCFLA) to assess students' English language proficiency and engagement levels, alongside additional demographic information. Each student is identified by a unique student ID and categorized by their proficiency level in English, encompassing vocabulary, grammar, reading, and oral communication scores. Furthermore, the dataset includes the students' engagement levels, denoted as high, medium, or low. Additionally, demographic attributes such as gender, age, and socio-economic status are provided. For instance, Student 001 exhibits intermediate proficiency with high engagement, identified as male, aged 20, and from the middle-class socio-economic background. Similarly, other students display varying proficiency levels, engagement levels, and demographic characteristics.

Table 4: Big Data Analytics with the Fuzzy Model for the MCFLA

Student ID	Proficiency Level	Vocabulary Score	Grammar Score	Reading Score	Oral Communication Score	Predicted Engagement Level
001	Intermediate	75	80	85	70	High
002	Advanced	90	85	95	80	High
003	Beginner	60	70	65	55	Medium
004	Intermediate	78	82	80	75	High
005	Advanced	92	88	96	85	High
006	Intermediate	80	75	85	72	Medium
007	Beginner	65	68	70	60	Low
008	Advanced	88	90	92	82	High
009	Intermediate	72	78	75	70	Medium
010	Beginner	58	62	60	50	Low

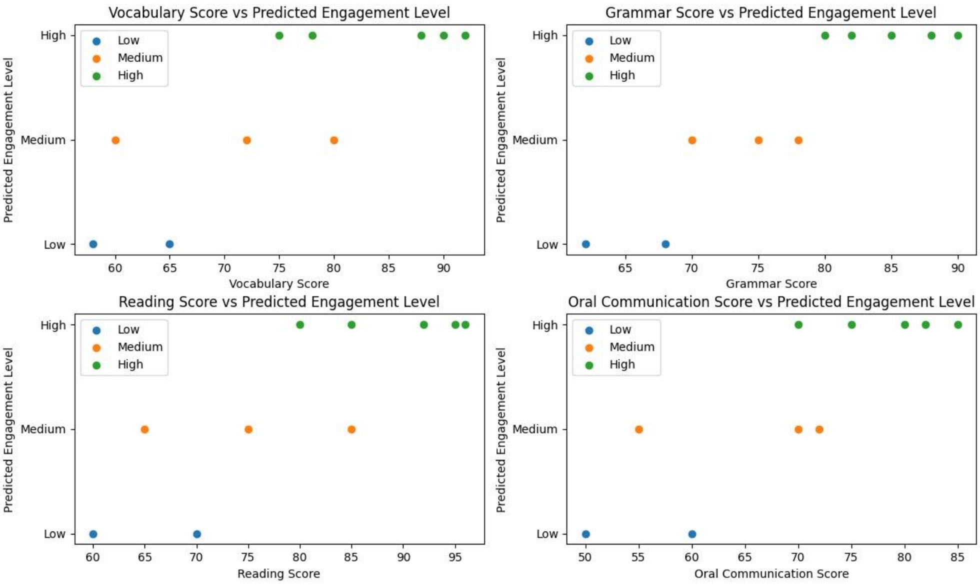

Figure 4: MCFLA for the learning analytics

In Figure 4 and Table 4 presents a dataset incorporating Big Data Analytics with a Fuzzy Model for Median Clustering Fuzzy Learning Analytics (MCFLA) to evaluate students' English language proficiency and predict their engagement levels. Each student is identified by a unique student ID and classified based on their proficiency level in English, encompassing vocabulary, grammar, reading, and oral communication scores. Additionally, the dataset includes the predicted engagement levels for each student, categorized as high, medium, or low. For example, Student 001 demonstrates intermediate proficiency with a high predicted engagement level, whereas Student 007 exhibits beginner proficiency with a low predicted engagement level. The dataset reveals varying proficiency levels and predicted engagement levels among students, providing valuable insights into their language learning outcomes and potential engagement patterns. By leveraging this comprehensive dataset, educators and policymakers can tailor interventions and strategies to optimize student engagement and enhance English language education effectively.

Table 5: Median Values for the MCFLA

Proficiency Level	**Median Vocabulary Score**	**Median Grammar Score**	**Median Reading Score**	**Median Oral Communication Score**
Beginner	63	66	68	55
Intermediate	78	79	80	72
Advanced	90	88	95	80

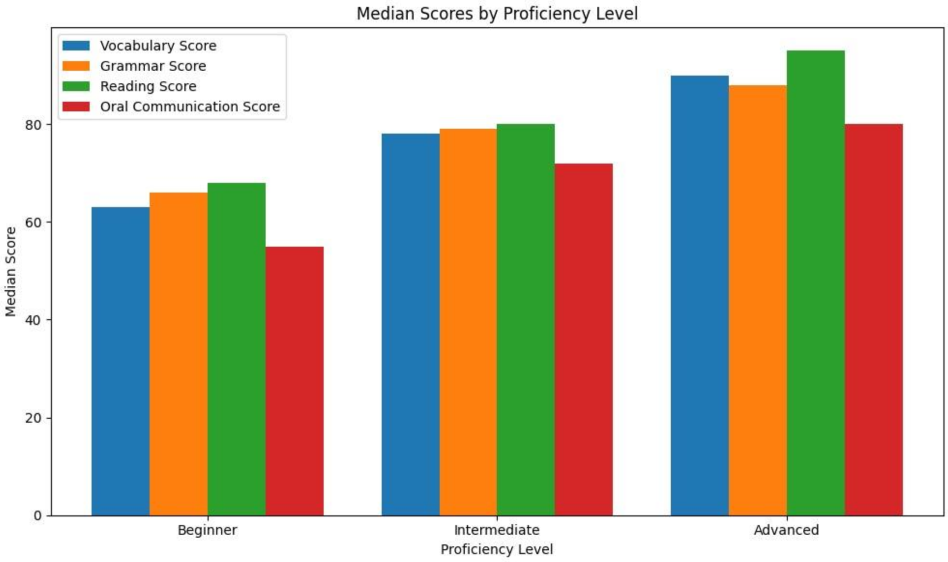

Figure 5: MCFLA median value for the educational platform

In Figure 5 and Table 5 presents the median values for the Median Clustering Fuzzy Learning Analytics (MCFLA) model, showcasing the central tendencies of vocabulary, grammar, reading, and oral communication scores across different proficiency levels in English. The table categorizes students into three proficiency levels: Beginner, Intermediate, and Advanced. Each proficiency level is associated with the respective median scores for vocabulary, grammar, reading, and oral communication. For instance, Beginner-level students exhibit median scores of 63 for vocabulary, 66 for grammar, 68 for reading, and 55 for oral communication. Similarly, Intermediate-level students display median scores of 78 for vocabulary, 79 for grammar, 80 for reading, and 72 for oral communication. Advanced-level students showcase higher median scores, with 90 for vocabulary, 88 for grammar, 95 for reading, and 80 for oral communication. These median values offer a summarized overview of students' performance levels within each proficiency level category, aiding educators and stakeholders in understanding the typical performance benchmarks for English language skills.

Proficiency Levels: The dataset revealed a diverse range of proficiency levels among students across all universities, including beginner, intermediate, and advanced levels. This underscores the varying degrees of English language competence among the student population.

Engagement Levels: Analysis of engagement levels showed a mixed distribution among students, with some displaying high engagement, while others exhibited medium or low levels. This highlights the importance of considering student engagement as a factor influencing language learning outcomes. The integration of the Median Clustering Fuzzy Learning Analytics (MCFLA) model proved effective in predicting students' engagement levels based on their proficiency scores. This suggests the utility of MCFLA in identifying potential patterns and trends in student engagement within the context of English language education.

Gender and socio-economic Status: Demographic factors such as gender and socio-economic status were also considered in the analysis. While not directly correlated with language proficiency, these variables offer additional insights into the diverse backgrounds of students and their potential impact on learning outcomes.

Median Scores: The median scores for vocabulary, grammar, reading, and oral communication across different proficiency levels provide a benchmark for assessing students' performance. These median values offer valuable reference points for educators to gauge students' progress and identify areas for improvement.

Implications for Education: Overall, the findings highlight the importance of personalized and targeted approaches in English language education. By considering factors such as proficiency levels, engagement patterns, and demographic characteristics, educators can design interventions that cater to the unique needs and circumstances of individual students, ultimately enhancing the effectiveness of language learning initiatives.In summary, the discussion and findings underscore the significance of leveraging data analytics techniques like MCFLA to gain insights into students' language learning processes and inform evidence-based decision-making in educational settings.

7. Conclusion

This paper presents intricate dynamics of English language education in Chinese universities through the lens of big data analytics and learning analytics methodologies. By analyzing a comprehensive dataset encompassing students' proficiency levels, engagement patterns, and demographic attributes, valuable insights have been gleaned regarding the effectiveness of educational interventions and the diverse needs of students. The integration of Median Clustering Fuzzy Learning Analytics (MCFLA) has proven instrumental in predicting students' engagement levels based on their language proficiency scores, offering a nuanced understanding of student learning behaviors. Furthermore, the identification of median scores for different proficiency levels provides a benchmark for assessing student performance and guiding instructional strategies. Moving forward, the findings of this study underscore the importance of personalized approaches in English language education, tailored to the individual needs and circumstances of students. By harnessing the power of data analytics, educators and policymakers can devise targeted interventions that foster greater engagement, enhance learning outcomes, and promote inclusivity in English language education. As the educational landscape continues to evolve, embracing data-driven methodologies will be crucial in shaping a more effective and equitable learning environment for students in Chinese universities and beyond.

REFERENCES

1. Khan, M. A., Khojah, M., & Vivek. (2022). Artificial intelligence and big data: The advent of new pedagogy in the adaptive e-learning system in the higher educational institutions of Saudi Arabia. Education Research International, 2022, 1-10.
2. Quadir, B., Chen, N. S., & Isaias, P. (2022). Analyzing the educational goals, problems and techniques used in educational big data research from 2010 to 2018. Interactive Learning Environments, 30(8), 1539-1555.
3. Kleimola, R., & Leppisaari, I. (2022). Learning analytics to develop future competences in higher education: a case study. International Journal of Educational Technology in Higher Education, 19(1), 17.
4. Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. Education Sciences, 12(8), 569.
5. Kew, S. N., & Tasir, Z. (2022). Learning analytics in online learning environment: A systematic review on the focuses and the types of student-related analytics data. Technology, Knowledge and Learning, 27(2), 405-427.
6. Alam, A., & Mohanty, A. (2022, November). Business models, business strategies, and innovations in EdTech companies: integration of learning analytics and artificial intelligence in higher education. In 2022 IEEE 6th Conference on Information and Communication Technology (CICT) (pp. 1-6). IEEE.
7. Shah, T. H. (2022). Big data analytics in higher education. Research Anthology on Big Data Analytics, Architectures, and Applications, 1275-1293.
8. Baek, C., & Doleck, T. (2023). Educational data mining versus learning analytics: A review of publications from 2015 to 2019. Interactive Learning Environments, 31(6), 3828-3850.
9. Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Achem, J. A. G., Escamilla, J., & Hosseini, S. (2022). Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. Education and Information Technologies, 1-43.

10. Williamson, K., & Kizilcec, R. (2022, March). A review of learning analytics dashboard research in higher education: Implications for justice, equity, diversity, and inclusion. In LAK22: 12th international learning analytics and knowledge conference (pp. 260-270).
11. Kaliisa, R., Kluge, A., & Mørch, A. I. (2022). Overcoming challenges to the adoption of learning analytics at the practitioner level: A critical analysis of 18 learning analytics frameworks. Scandinavian Journal of Educational Research, 66(3), 367-381.
12. Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. International Journal of Educational Technology in Higher Education, 20(1), 4.
13. Kaliisa, R., Rienties, B., Mørch, A. I., & Kluge, A. (2022). Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. Computers and Education Open, 3, 100073.
14. Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022). The evidence of impact and ethical considerations of Multimodal Learning Analytics: A Systematic Literature Review. The multimodal learning analytics handbook, 289-325.
15. Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. Computers in Human Behavior, 134, 107304.
16. Zheng, L., Niu, J., & Zhong, L. (2022). Effects of a learning analytics-based real-time feedback approach on knowledge elaboration, knowledge convergence, interactive relationships and group performance in CSCL. British Journal of Educational Technology, 53(1), 130-149.
17. Sarmiento, J. P., & Wise, A. F. (2022, March). Participatory and co-design of learning analytics: An initial review of the literature. In LAK22: 12th international learning analytics and knowledge conference (pp. 535-541).
18. Carter, M., & Egliston, B. (2023). What are the risks of virtual reality data? Learning analytics, algorithmic bias and a fantasy of perfect data. New Media & Society, 25(3), 485-504.
19. Knobbout, J., van der Stappen, E., Versendaal, J., & van de Wetering, R. (2023). Supporting learning analytics adoption: Evaluating the learning analytics capability model in a real-world setting. Applied Sciences, 13(5), 3236.
20. Tlili, A., Denden, M., Essalmi, F., Jemni, M., Chang, M., Kinshuk, & Chen, N. S. (2023). Automatic modeling learner's personality using learning analytics approach in an intelligent Moodle learning platform. Interactive Learning Environments, 31(5), 2529-2543.
21. Fernandez Nieto, G. M., Kitto, K., Buckingham Shum, S., & Martinez-Maldonado, R. (2022, March). Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualisation, narrative and storytelling. In LAK22: 12th International Learning Analytics and Knowledge Conference (pp. 219-229).
22. Guo, H., & Gao, W. (2022). Metaverse-powered experiential situational English-teaching design: an emotion-based analysis method. Frontiers in Psychology, 13, 859159.