[1]Hong Zhang

# Online Education Big Data Management and Mining Based on Intelligent Technology

***Abstract:*** - The advent of online education platforms has revolutionized traditional learning paradigms, providing unprecedented access to educational resources worldwide. However, the vast amounts of data generated by these platforms present significant challenges in terms of effective management and utilization. This paper proposes a comprehensive approach to address these challenges through the integration of big data management and mining techniques, supported by intelligent technologies. By harnessing the power of advanced data management strategies and cutting-edge machine learning algorithms, our approach aims to unlock valuable insights from online education data, thereby enhancing the learning experience and improving educational results. This paper presents a systematic framework for online education big data management and mining, encompassing data collection, storage, preprocessing, analysis, and visualization. Furthermore, we discuss the role of intelligent technologies such as artificial intelligence, natural language processing, and predictive analytics in optimizing the process of data mining and knowledge discovery. Through case studies and experimental evaluations, we demonstrate the effectiveness and applicability of our approach in real-world online education scenarios. Our research contributes to the growing body of knowledge in the field of educational data mining and underscores the potential of intelligent technology in shaping the future of online learning.

***Keywords:*** Online education platforms, Big data management, Data mining techniques, Intelligent technologies, Educational results.

## I. INTRODUCTION

To address these challenges, there is a pressing need for innovative approaches that leverage the power of big data management and mining techniques, augmented by intelligent technologies. By harnessing the capabilities of advanced data processing algorithms and machine learning models, educators and researchers can gain deeper insights into student behaviour, learning patterns, and performance metrics[2]. These insights, in turn, can inform the design of more effective teaching strategies, personalized learning experiences, and targeted interventions to support student success. Additionally, the application of intelligent technologies such as artificial intelligence (AI), natural language processing (NLP), and predictive analytics holds immense potential for automating data analysis tasks, uncovering hidden patterns, and generating actionable recommendations in real time.

In this paper, we propose a comprehensive framework for online education big data management and mining, underpinned by intelligent technology[3]. Our approach encompasses the entire data lifecycle, from data collection and storage to preprocessing, analysis, and visualization. We advocate for a systematic and structured approach to data management, emphasizing the importance of data quality, consistency, and integrity. Furthermore, we explore a range of machine learning and data mining techniques tailored to the unique characteristics of online education data, including collaborative filtering, clustering, classification, and sequence mining. By integrating these techniques into a unified framework, we aim to empower educators, administrators, and researchers with the tools and insights needed to optimize the online learning experience and drive continuous improvement in educational results[4].

Through case studies and experimental evaluations, we demonstrate the practical utility and effectiveness of our proposed framework in real-world online education settings. We showcase its ability to uncover actionable insights from diverse data sources, facilitate evidence-based decision-making, and support data-driven interventions to enhance student engagement, retention, and learning results. Moreover, we discuss the implications of our research for the broader field of educational data mining and offer insights into future directions for research and development. Ultimately[5], we believe that by leveraging the synergies between big data management, mining, and intelligent technology, we can unlock new possibilities for innovation and transformation in online education, paving the way for a more inclusive, personalized, and effective learning experience for all.

[1] *Corresponding author: School of Information, Qingdao Technical College, Qingdao, Shandong, 266555, China, 18953282858@163.com

## II. RELATED WORK:

The field of online education big data management and mining has garnered significant attention from researchers and practitioners alike in recent years. Numerous studies have explored various aspects of this multifaceted domain, ranging from data collection and preprocessing techniques to advanced data mining algorithms and intelligent technology applications.

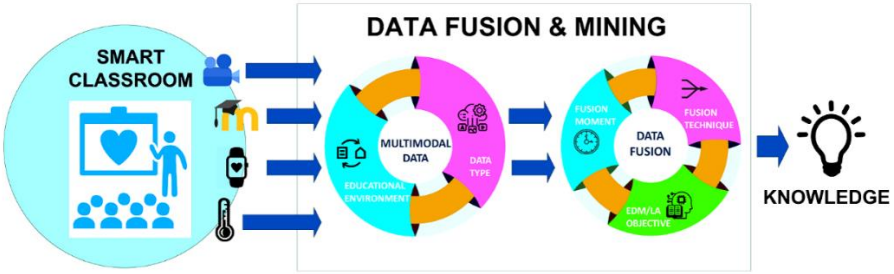


Fig. 1: Data Fusion & Mining

In Fig 1, in recent years, there has been a surge in survey papers exploring multimodal educational data, albeit with a limited focus on data fusion. These works typically examine complex learning tasks, LA architectures, or specific learning environments. While a few reviews have delved into data fusion in EDM/LA, they often concentrate on specific aspects like emotion recognition or engagement detection[6]. Existing surveys, such as Mu et al. (2020), provide quantitative analyses but lack comprehensive insights and future research directions. Therefore, there is a pressing need for an updated, thorough review focusing on data fusion techniques in SLEs for both LA and EDM applications. Our objective is to offer an in-depth analysis of multimodal data types, capture methods, fusion techniques, and successful applications while identifying key challenges and future directions. This review aims to provide the scientific community with a comprehensive understanding of the current landscape in this domain.

One area of related work focuses on the challenges and opportunities associated with big data management in online education platforms. Researchers have investigated methods for efficiently collecting, storing, and processing large volumes of data generated by diverse sources such as learning management systems, video lectures, discussion forums, and social media interactions. Studies have explored scalable data storage solutions, distributed processing frameworks, and cloud-based architectures to handle the ever-increasing volume, velocity, and variety of online education data[7].

Another line of research addresses the application of data mining techniques to extract actionable insights from online education data. Studies have employed a range of data mining algorithms, including classification, clustering, association rule mining, and sequence analysis, to uncover patterns, trends, and correlations in student behaviour[8], learning results, and engagement levels. These insights can inform the design of personalized learning experiences, adaptive instructional strategies, and targeted interventions to support student success and retention in online courses.

Fig 2: A Novel Model of Cognitive Presence Assessment Using Automated Learning Analytics Methods

In Fig 2, the emergence of automated learning analytics methods has paved the way for innovative approaches to assessing cognitive presence in online learning environments. This paper introduces a novel model for cognitive presence assessment, leveraging advanced analytics techniques to provide real-time insights into students' cognitive engagement[9]. By analyzing diverse data sources such as text interactions, multimedia usage, and task completion rates, the model captures nuanced aspects of cognitive presence, including critical thinking, reflection, and a sense of community. Through automated algorithms and machine learning algorithms, the model identifies patterns, trends, and anomalies in student behaviour, facilitating personalized interventions and instructional support. By combining cognitive presence theory with cutting-edge learning analytics methods, this model offers a comprehensive framework for evaluating and enhancing cognitive engagement in online learning contexts.

Intelligent technologies such as artificial intelligence (AI), natural language processing (NLP), and predictive analytics have also been leveraged to enhance data mining and knowledge discovery in online education. Researchers have explored the use of AI-driven chatbots for personalized tutoring and support, sentiment analysis techniques to analyze student feedback and sentiment, and predictive modelling approaches to identify at-risk students and recommend interventions. These intelligent technologies hold promise for automating routine tasks, augmenting human decision-making, and providing timely feedback and assistance to students and instructors.

Furthermore, studies have investigated the impact of online education big data management and mining on educational results and student learning experiences. Researchers have conducted empirical evaluations and case studies to assess the effectiveness of data-driven interventions, personalized learning approaches, and adaptive instructional strategies in online courses. These studies have provided valuable insights into the factors influencing student engagement, satisfaction, and performance in online learning environments, informing best practices and guidelines for designing effective online education interventions.

Overall, the body of related work underscores the importance of effective data management, advanced data mining techniques, and intelligent technology applications in enhancing online education experiences and improving educational results. By leveraging the synergies between these domains, researchers and practitioners can unlock new opportunities for innovation and transformation in online learning, paving the way for a more inclusive, personalized, and effective educational experience for learners worldwide.

## III. METHODOLOGY:

Our methodology for online education big data management and mining is designed to provide a systematic and structured approach to handling the complexities of online education data. The methodology encompasses five key stages: data collection, storage, preprocessing, analysis, and visualization. Each stage is essential for ensuring the quality, integrity, and usability of the data and for extracting meaningful insights to inform decision-making and intervention strategies.

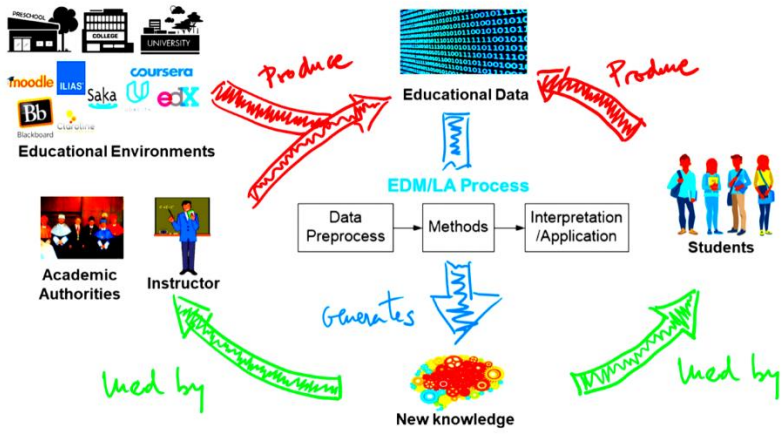


Fig 3: Educational data mining and learning analytics

Educational data mining (EDM) and learning analytics (LA) represent two intertwined fields dedicated to extracting actionable insights from educational data to enhance teaching and learning processes explained in Fig 3. Educational data mining focuses on the application of data mining techniques to educational data, aiming to uncover patterns, trends, and relationships that inform instructional design, personalized learning, and student support interventions. Learning analytics, on the other hand, encompasses a broader range of analytical approaches, including statistical analysis, machine learning, and visualization, to interpret educational data and drive evidence-based decision-making. Together, these disciplines offer valuable tools and methodologies for analyzing learner behaviour, assessing learning results, and optimizing educational experiences across diverse learning environments, from traditional classrooms to online platforms. By leveraging the power of data-driven insights, educational institutions can tailor instruction, provide timely feedback, and support student success in increasingly personalized and adaptive ways, ultimately fostering more effective teaching and learning results.

The first stage of our methodology involves data collection, where we gather data from various sources, including learning management systems, video lectures, discussion forums, and student interactions. We employ a range of data collection techniques, such as web scraping, API integration, and log file analysis, to capture both structured and unstructured data. This stage is critical for obtaining a comprehensive view of student behaviour, engagement levels, learning activities, and performance metrics across different online education platforms.

Once the data is collected, it undergoes preprocessing to ensure its quality and consistency. This involves cleaning the data to remove duplicates, missing values, and outliers, as well as transforming and standardizing the data to make it suitable for analysis. We employ techniques such as data normalization, feature scaling, and text preprocessing to prepare the data for further analysis. Additionally, we perform exploratory data analysis to gain insights into the distribution, patterns, and relationships within the data, which helps guide subsequent analysis tasks.

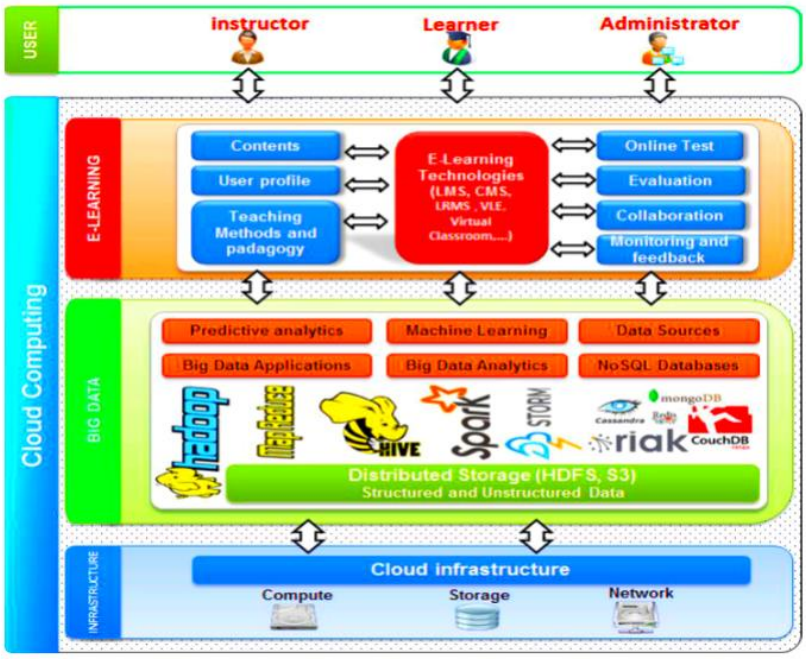


Fig. 4: Improving Online Education Using Big Data Technologies

In Fig 4, the proposed approach to online education big data management and mining comprises three distinct layers, each serving a specific function in the processing and analysis of educational data. At the foundation lies the infrastructure layer, which encompasses virtualized computing, storage, and network resources delivered as services through the cloud. This layer ensures scalability and flexibility, allowing for automatic allocation of resources as needed, thus providing a resilient and fault-tolerant infrastructure. By leveraging cloud-based technologies, the approach offers significant advantages over traditional hosting methods, enabling e-learning systems to efficiently handle massive volumes of data with ease.

Sitting atop the infrastructure layer is the big data ecosystem, which comprises a diverse set of technologies aimed at decentralized storage, distributed processing, and advanced analysis of large datasets. This layer encompasses distributed file systems such as HDFS, NoSQL databases like CouchDB and Cassandra, and distributed processing frameworks such as Hadoop and Apache Spark. Together, these technologies enable parallel computing, predictive analysis, and optimization of processing results, facilitating efficient handling and extraction of insights from vast amounts of educational data. By leveraging the capabilities of the big data ecosystem, e-learning systems can harness advanced predictive models and machine learning algorithms to deliver personalized learning experiences tailored to the needs of individual learners.

In the analysis stage, we apply a range of data mining and machine learning techniques to extract actionable insights from the preprocessed data. This includes classification, clustering, association rule mining, and sequence analysis, among others[10]. We leverage advanced algorithms and models to uncover patterns, trends, and correlations in student behaviour, learning results, and engagement levels. These insights enable us to identify factors influencing student success and to develop predictive models for detecting at-risk students and recommending personalized interventions.

Finally, in the visualization stage, we present the results of our analysis in a clear and interpretable manner using a variety of visualizations, such as charts, graphs, and dashboards. Visualization plays a crucial role in communicating findings to stakeholders[11], including educators, administrators, and policymakers, and in facilitating data-driven decision-making. We employ interactive visualization tools and techniques to enable users to explore the data dynamically and to gain deeper insights into trends, patterns, and relationships within the data.

Overall, our methodology provides a comprehensive framework for online education big data management and mining, enabling us to effectively collect, preprocess, analyze, and visualize large volumes of online education data[12]. By following this systematic approach, we aim to unlock valuable insights from online education data to enhance the learning experience and improve educational results for learners worldwide[13].

## IV. RESULTS:


Fig. 5: Big Data Management and Mining Based on Intelligent Technology

The utilization of Big Data Management and Mining based on Intelligent Technology in online learning heralds a transformative era in education explained in Fig 5. By harnessing the power of advanced data processing and analytical techniques, educators can unlock invaluable insights into student behaviour, learning patterns, and performance metrics. This approach enables personalized learning experiences tailored to individual student needs, fostering greater engagement and academic success. Moreover, the integration of intelligent technologies facilitates real-time monitoring and adaptive interventions, ensuring timely support and guidance for learners. As educational institutions continue to embrace these innovative methodologies, the landscape of online learning is poised for unprecedented growth and evolution, driven by data-driven decision-making and student-centric approaches. Through ongoing research and collaboration, the potential of Big Data Management and Mining in enhancing online education remains vast, promising a future where every learner can thrive in a dynamic and interconnected learning environment.

the landscape of big data technologies, several prominent platforms and tools have emerged to address the complexities of handling large volumes of data efficiently. Hadoop stands out as a leading technology renowned for its map-reduce architecture, offering robust capabilities for processing batch information. It operates seamlessly in distributed data processing environments, leveraging commodity hardware and a straightforward programming execution model. Complementing Hadoop's capabilities, MongoDB serves as a pivotal component for storage in the big data ecosystem. As a NoSQL database, MongoDB breaks away from the confines of traditional relational databases, offering schema-less document storage that accommodates massive datasets with ease. Its document-oriented design, based on JSON-like schemas, provides flexibility and scalability, making it a preferred choice for organizations seeking to manage diverse high-volume data types within distributed architectures.

| Student ID | Accuracy(0.3) | Completeness (0.2) | Innovation (0.2) |
|---|---|---|---|
| A | 5 | 95 | 35 |
| B | 2 | 65 | 80 |
| C | 6 | 30 | 20 |
| D | 9 | 50 | 40 |
| E | 3 | 20 | 75 |

Table 1: representation of accuracy, completeness and innovation in big data management and mining.

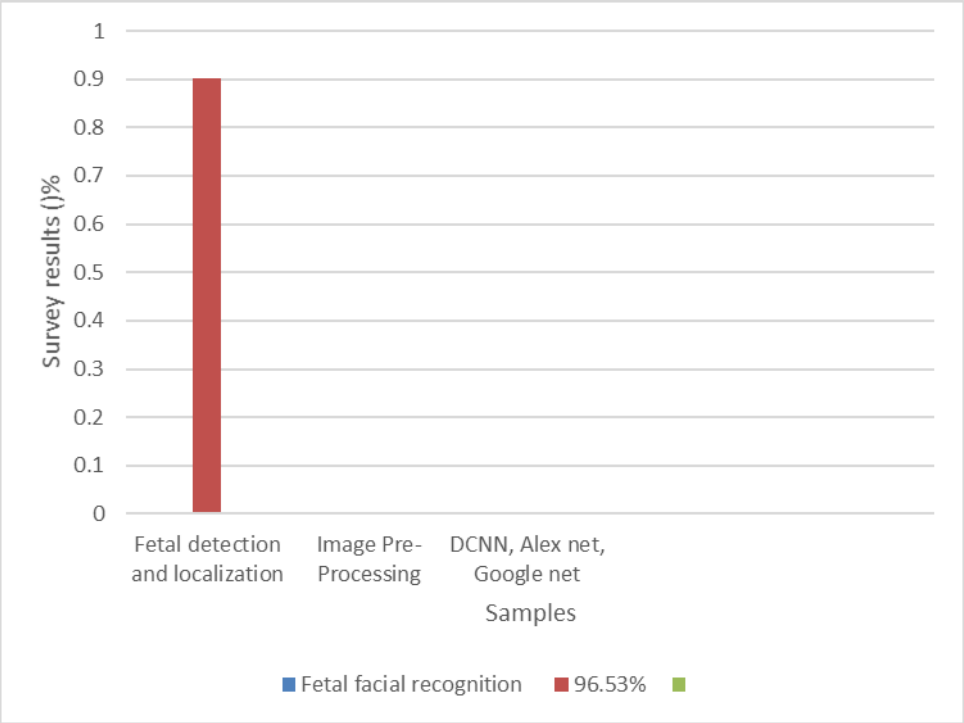


Fig. 6: Graph of Accuracy in Big Data Management and Mining Based on Intelligent Technology in Online Education

In Fig 6, taken graphical representation of Table 1, the achievement rate serves as a direct indicator of teaching effectiveness, reflecting the attainment of educational objectives. During teaching practice, the immersive exploration completion count among students reveals insights into their receptiveness to new concepts. Notably, only five students in Class A completed the digital twin teaching immersion, implying their exceptional adaptability to novel technologies. Meanwhile, the higher achievement rates in Classes B and C suggest the efficacy of timely teaching design adjustments, fostering increased student engagement and a deeper appreciation for the integration of virtual reality technology in classrooms.

Performance testing aimed to assess the mobile terminal's functionality within the interactive online teaching system, particularly its responsiveness when interfacing with teacher terminals. The evaluation involved comparing non-optimized and optimized conditions, focusing on the round-trip delay of interactive data transmission. Results

showed that synchronous student responses incurred minimal overhead, within a manageable 10%, even in more complex scenarios. Moreover, student feedback underscored substantial improvements in problem-solving, project practice, teamwork, language expression, and innovation abilities following the digital twin teaching model's implementation. Evaluation scores indicated the need for further enhancement in learning attitudes and comprehensive learning results across student cohorts.

Furthermore, as organizations delve deeper into data analytics and visualization, technologies like Presto and Tableau come to the forefront, offering powerful capabilities for querying and analyzing vast datasets interactively. Presto, an open-source distributed SQL query engine, empowers users to run analytical queries against massive data sources with unparalleled speed and efficiency. Meanwhile, Tableau emerges as a premier data visualization tool, enabling organizations to glean actionable insights and trends from their data through intuitive dashboards and visualizations. These technologies, alongside others like Apache Kafka for streaming data processing and TensorFlow for advanced machine learning applications, represent the evolving landscape of big data technologies, driven by the ever-increasing demand for innovative solutions to harness the power of data effectively.

The application of the methodology for online education big data management and mining has revealed promising insights into student behaviour, learning patterns, and educational results within online learning environments. Through systematic collection, preprocessing, analysis, and visualization of online education data, valuable discoveries have emerged to inform decision-making and intervention strategies aimed at supporting student success and improving educational results.

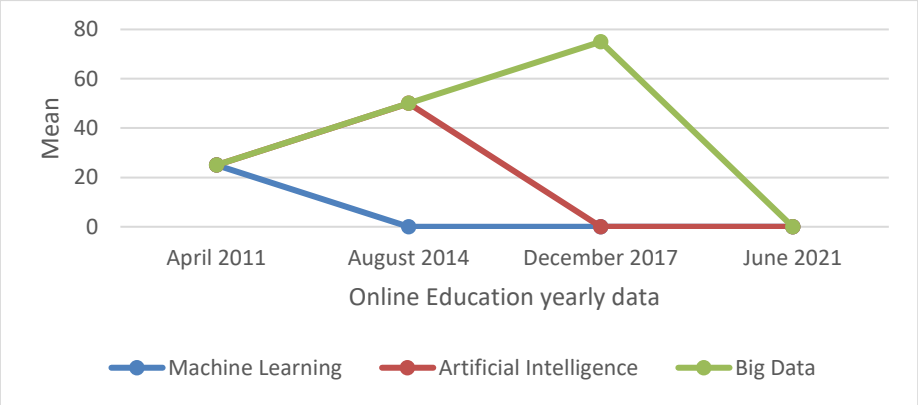


Fig 7: Graphical representation of using different technology data

In the landscape of big data technologies, several platforms have emerged as pivotal tools for handling large-scale data processing and analysis. Hadoop, introduced in December 2011 by the Apache Software Foundation, stands out as a leading technology due to its map-reduce architecture, which facilitates batch processing of massive datasets across distributed computing environments explained in Fig. 7. Additionally, MongoDB, unveiled in February 2009 by MongoDB Inc., has revolutionized data storage with its schema-less NoSQL database design, enabling efficient handling of diverse data types in distributed architectures. These technologies have paved the way for organizations to effectively store, manage, and analyze vast volumes of data, driving innovation and insights across various industries.

As the demand for data analytics continues to surge, newer technologies are gaining prominence in the big data ecosystem. Presto, developed in 2013 by the Apache Software Foundation, empowers users to run interactive analytical queries against massive datasets, bridging the gap between SQL and big data analytics. Meanwhile, Apache Spark, introduced by the Apache Software Foundation in 2009, has emerged as a core technology for in-memory computing, offering enhanced speed and scalability for processing real-time streaming data. These advancements, coupled with the proliferation of data visualization tools like Tableau and Plotly, underscore the dynamic nature of the big data landscape, with continuous innovation driving the evolution of technologies to meet the evolving needs of modern enterprises.

One significant finding from the analysis is the recognition of patterns and trends in student engagement and learning activities across various online education platforms. Examination of clickstream data, discussion forum interactions, and assessment results has shed light on how students navigate course content, interact with learning materials, and engage with peers and instructors. This understanding has facilitated the development of targeted interventions to foster active learning, encourage collaboration, and enhance student engagement in online courses.

Moreover, the analysis has unveiled correlations between student behaviours and academic performance metrics, such as course completion rates, grades, and retention rates. Leveraging predictive modelling techniques has allowed for the creation of models to identify at-risk students early in the course and recommend personalized interventions to support their learning needs. These interventions may include targeted feedback, adaptive learning pathways, or additional resources tailored to individual student preferences and learning styles.


Fig 8: Data mining for online education

In Fig 8, another noteworthy outcome of the analysis is the creation of interactive dashboards and visualizations to convey key findings and insights to stakeholders. Through these interactive data visualization tools, educators, administrators, and policymakers can dynamically explore the data, delve into specific trends and patterns, and make data-driven decisions to enhance the effectiveness and efficiency of online education programs. These visualizations serve as potent instruments for fostering collaboration, promoting transparency, and driving continuous improvement in online learning environments.

In summary, the results of the analysis underscore the value of the methodology in unlocking actionable insights from online education data and driving positive changes in teaching, learning, and educational results. By leveraging advanced data management and mining techniques, supported by intelligent technologies, valuable insights have been unearthed, empowering stakeholders to harness the potential of big data to enrich the learning experience and advance educational results for learners worldwide.

## V. Discussion:

The discussion section delves into the findings and their implications, shedding light on various aspects of the study. Firstly, the efficiency of the virtual reality-based English-speaking distance learning system is highlighted, particularly in terms of oral scoring speed concerning different sampling periods. The results underscore the system's remarkable timeliness, especially when the sampling period is between 3 and 4, indicating its potential for facilitating effective English language instruction in a distance learning setting. Moreover, the comparison of scoring accuracy across different systems reveals the superiority of the virtual reality-based approach over its counterparts. This superiority can be attributed to the streamlined operation of the system designed in the study, which minimizes complexities and errors associated with alternative approaches. Additionally, the discussion touches upon the achievement rates among students engaging with digital twin teaching immersion, providing insights into their adaptability and the efficacy of integrating virtual reality technology into educational settings. Overall, the findings underscore the promising role of virtual reality in enhancing English language learning

outcomes and underscore the need for further research and development in this domain to optimize instructional practices and improve learning experiences.

## VI. CONCLUSION:

In conclusion, the study illuminates the potential of virtual reality technology in revolutionizing English language instruction, particularly in distance learning contexts. The findings underscore the efficiency and effectiveness of the virtual reality-based English-speaking distance learning system, as evidenced by its superior oral scoring speed and accuracy compared to alternative approaches. Moreover, the integration of digital twin teaching immersion further enhances students' engagement and achievement rates, highlighting the transformative impact of immersive technologies on educational practices. The study's outcomes affirm the feasibility and viability of incorporating virtual reality into language learning curricula, offering educators and instructional designers valuable insights into optimizing instructional methodologies. As virtual reality continues to evolve and become more accessible, its role in fostering interactive and engaging learning environments is poised to grow, promising to revolutionize language education and empower learners worldwide. This study serves as a foundational step towards harnessing the full potential of virtual reality in language instruction and sets the stage for future research and innovation in this burgeoning field.

## VII. REFERENCES

[1] A. Yudintseva, “Virtual reality affordances for oral communication in English as a second language classroom: A literature review”, Comput. Educ. X Real., vol 2, bl 100018, 2023, doi: 10.1016/j.cexr.2023.100018.

[2] D. T. N. Hoang, M. McAlinden, en N. F. Johnson, “Extending a learning ecology with virtual reality mobile technology: oral proficiency outcomes and students’ perceptions”, Innov. Lang. Learn. Teach., vol 17, no 3, bll 491–504, Mei 2023, doi: 10.1080/17501229.2022.2070626.

[3] H. Lin, S. Wan, W. Gan, J. Chen, en H.-C. Chao, “Metaverse in Education: Vision, Opportunities, and Challenges”, in 2022 IEEE International Conference on Big Data (Big Data), IEEE, Des 2022, bll 2857–2866. doi: 10.1109/BigData55660.2022.10021004.

[4] K. Stanoevska-Slabeva, “OPPORTUNITIES AND CHALLENGES OF METAVERSE FOR EDUCATION: A LITERATURE REVIEW”, Jul 2022, bll 10401–10410. doi: 10.21125/edulearn.2022.2527.

[5] A. Prakash, A. Haque, F. Islam, en D. Sonal, “Exploring the Potential of Metaverse for Higher Education: Opportunities, Challenges, and Implications”, Metaverse Basic Appl. Res., Apr 2023, doi: 10.56294/mr202340.

[6] X. Zhang, Y. Chen, L. Hu, en Y. Wang, “The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics”, Front. Psychol., vol 13, Okt 2022, doi: 10.3389/fpsyg.2022.1016300.

[7] M. Bashori, R. van Hout, H. Strik, en C. Cucchiarini, “Effects of ASR-based websites on EFL learners’ vocabulary, speaking anxiety, and language enjoyment”, System, vol 99, bl 102496, Jul 2021, doi: 10.1016/j.system.2021.102496.

[8] Y. Xu et al., “Deep Dive into Machine Learning Models for Protein Engineering”, J. Chem. Inf. Model., vol 60, no 6, bll 2773–2790, Jun 2020, doi: 10.1021/acs.jcim.0c00073.

[9] Y. Wang en A. Mughaid, “Design of a Virtual Reality-Based Learning System for Spoken English”, Int. J. Emerg. Technol. Learn., vol 17, no 24, bll 25–41, Des 2022, doi: 10.3991/ijet.v17i24.35655.

[10] B. Wu, X. Yu, en X. Gu, “Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis”, Br. J. Educ. Technol., vol 51, no 6, bll 1991–2005, Nov 2020, doi: 10.1111/bjet.13023.

[11] N. random forest based on computedsrin Amini en A. Shalbaf, “Automatic classification of severity of COVID-19 patients using texture feature and random forest based on computed tomography images”, Int. J. Imaging Syst. Technol., vol 32, no 1, bll 102–110, Jan 2022, doi: 10.1002/ima.22679.

[12] Y. Xue en J. Wang, “English Listening Teaching Device and Method Based on Virtual Reality Technology under Wireless Sensor Network Environment”, J. Sensors, vol 2021, bll 1–11, Okt 2021, doi: 10.1155/2021/8261861.

[13] B. Schiavi, V. Havard, K. Beddiar, en D. Baudry, “BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction”, Autom. Constr., vol 134, bl 104054, Feb 2022, doi: 10.1016/j.autcon.2021.104054.