IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 May 2024, accepted 18 May 2024, date of publication 21 May 2024, date of current version 31 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403758

RESEARCH ARTICLE

Application of Artificial Intelligence in Online Education: Influence of Student Participation on Academic Retention in Virtual Courses

WILLIAM VILLEGAS-CH[1], **(Member, IEEE), JOSELIN GARCÍA-ORTIZ**[1], **AND SANTIAGO SÁNCHEZ-VITERI**[2]

[1]Escuela de Ingeniería en Ciberseguridad, FICA, Universidad de Las Américas, Quito 170125, Ecuador
[2]Departamento de Sistemas, Universidad Internacional del Ecuador, Quito 170411, Ecuador

Corresponding author: William Villegas-Ch (william.villegas@udla.edu.ec)

ABSTRACT The accelerated growth of online education in recent decades has made it a trendy educational option. Despite this, significant challenges persist regarding student retention and academic performance in these courses. Addressing these challenges requires active student participation, although understanding it effectively is complex. This study focuses on student engagement in online education environments, exploring its relationship with retention and academic performance. Data was collected across multiple semesters to analyze student engagement in online educational environments. Participation patterns, temporal trends, and crucial factors affecting participation were examined using exploratory analysis and forecasting models, such as ARIMA and Prophet. The results revealed several patterns, including an initial increase in activity at the course's beginning and a gradual decrease over time. Factors such as course length and peer interaction influenced participation significantly. These findings underscore the importance of developing specific pedagogical strategies for online education, simultaneously addressing students' unique challenges in this environment. In summary, this study contributes to knowledge in online education by providing essential information to understand and improve student engagement.

INDEX TERMS Student participation, academic retention, online courses.

I. INTRODUCTION

The exponential increase in online education in recent decades has made it an increasingly popular and accessible educational option. Although traditional institutions are expanding their digital offerings and specialized online universities are emerging, online education faces unique challenges, especially in student retention and academic performance [1]. Active and meaningful student participation is critical to success in any educational environment. In face-to-face classrooms, instructors can assess participation through direct interactions and discussions, but in the online environment, where there is no physical presence, evaluating and encouraging participation be-comes a complex challenge [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

The importance of student engagement in online education cannot be overstated. Engaged students tend to have higher motivation levels, exhibit better time management skills, and are likelier to persist and succeed in their courses. Conversely, a lack of engagement can lower retention rates, academic performance, and dropout risk. Therefore, understanding and improving student engagement in online classes have become critical objectives for educational institutions and researchers [3].

This work explores the landscape of student engagement in online education and its intimate connection to student retention and academic performance. It aims to unravel the factors influencing student engagement in the online learning environment, exploring how course length, peer interaction, and engagement patterns impact a student's educational journey. To achieve these goals, the basis of the analysis will be comprehensive data collected from a system designed to track student participation in an online academic environment over multiple semesters [4].

For this, two years of student participation in an online computer science program, which is part of the undergraduate educational offering of a recognized university, were analyzed. This official program offers a degree in Computer Science and is designed for students with prior knowledge in mathematics and programming. The goal is to provide a comprehensive and rigorous education, preparing students for professional challenges in the technological field. The diversity of students' academic backgrounds and previous online learning experiences enriches the analysis, allowing us to explore how different educational trajectories impact participation and engagement in a virtual learning environment.

The methodology used in this research combines exploratory analysis with the application of forecasting models. These models, including autoregressive integrated moving average (ARIMA), Prophet, and clustering techniques such as DBSCAN and K-means, are leveraged to decipher participation patterns, temporal trends, and critical determinants of student engagement. In doing so, we aim to provide valuable insights into the dynamics of student engagement in online courses and ultimately shed light on how institutions can improve the online learning experience for their students [5]. Additionally, these methods allow for identifying trends, seasonality, and grouping of students with similar behaviors, facilitating detailed understanding and accurate predictions of student engagement over time.

Online education poses unique challenges compared to traditional in-person instruction. Students in virtual classrooms often face distractions in their home environments, work commitments, personal demands, and the need to maintain high levels of motivation and self-discipline. These external factors can significantly influence your participation in course material, class discussions, and assignments. Therefore, understanding the interplay between these challenges and student engagement is vital to improving the quality and effectiveness of online education.

One of the critical contributions of this study lies in its evaluation of forecasting models, specifically ARIMA and Prophet, for predicting student engagement in online courses [6]. These models are evaluated for their ability to anticipate student engagement, allowing educational institutions to identify students at risk of low concentration early in the system [7]. Early identification will enable institutions to intervene proactively, providing targeted support to struggling students and improving their learning experience. The application of forecasting models in the context of online education represents a promising avenue for improving retention rates and student out-comes [8].

Additionally, this research contributes to the broader body of knowledge on online education and tracking student engagement. It aligns with ongoing efforts to improve the quality and inclusion of online education, ensuring that it continues to meet the academic expectations and needs of a diverse and dynamic student population. As the educational landscape evolves, particularly considering recent global events, adapting and innovating educational strategies is imperative to foster meaningful and compelling learning experiences.

II. MATERIALS AND METHODS

This work presents a detailed study of the implementation and evaluation landscape of the student participation monitoring system in the online educational environment. The virtual platform and its specific configuration, the activity recording tools used to capture student data, advanced data analysis, and machine learning techniques were analyzed for this. In addition, a personalized recommendation system has been implemented to improve collaboration between students. This approach allows for a complete understanding of the underlying methodology of the study.

Due to confidentiality restrictions and privacy agreements with the participating institution, we are not authorized to reveal specific information such as the name of the institution, the countries where the courses are offered, the country of origin of the participants, or the degrees involved. These restrictions protect the privacy of participants and the integrity of the institution.

A. REVIEW OF SIMILAR WORKS

The choice of journals and articles cited in the introduction and review of previous studies was based on a selective and strategic approach. We sought to identify work that would provide a solid and relevant foundation for the research presented in this article. Selection criteria included direct relevance to the topic of student engagement in online learning environments, as well as the quality and authority of the sources. Studies and publications that offered empirical evidence, recent research, and significant results in online education were prioritized.

Student engagement in online learning environments has been the subject of research for several decades. As online education has grown in importance, it has become essential to understand how students interact and participate in these environments. The literature review reveals that while significant progress has been made, most studies still face challenges quantifying and improving student engagement [9].

Recent research has highlighted the critical role of active student participation as a crucial factor for success in online education [10]. Studies have shown that students who actively participate in online discussions, collaborate on group projects, and participate in interactive activities are more likely to achieve satisfactory academic performance [11]. Additionally, it has been argued that student engagement may be a key indicator of student satisfaction and retention in online courses [12].

Learning analytics has developed significantly over the last decade and has contributed substantially to understanding online student engagement [13], [14]. Researchers have employed data analytics and machine learning techniques to examine patterns of student engagement and performance in virtual environments [15]. An exciting advance has been the development of recommendation systems based on learning analytics, which use advanced algorithms to analyze students' browsing and participation behaviors, offering personalized recommendations for resources and activities [16].

Despite these advances in the literature, it has been recognized that the practical implementation of these technologies in educational settings remains limited [17]. This study addresses such limitations through the practical application of a tracking system in an academic environment, illustrating how integrating activity recording tools, data analysis, and recommendation systems can significantly improve student participation and the quality of the online learning experience [18], [19]. By comparing the findings of our study with the existing literature, we highlight the unique contributions of our research, especially in terms of practical applicability and technological improvements in monitoring and analyzing student participation [20].

B. CONCEPTS USED

Several concepts have been employed to develop the methodology contributing to this study and each applicable aspect of the application environment. These concepts are utilized in this study to provide a clearer understanding of the technologies and strategies involved:

- Student engagement is central in online education, encompassing how students interact and actively engage in the learning processes within a virtual environment. This entails participation in discussion forums, collaboration on group projects, submission of homework, and the degree of involvement with course content. Student engagement is an essential indicator of students' interaction and engagement levels with the learning materials [21].
- Machine learning, a branch of AI, focuses on developing algorithms that enable computers to learn and enhance their performance in specific tasks through experience and data. This study employs machine learning techniques to analyze student engagement data and forecast behavioral patterns. This facilitates the provision of personalized recommendations to students and educators.
- A recommender system utilizes algorithms to analyze user behavior and preferences, delivering personalized recommendations. In the context of online education, these systems can propose learning resources, supplementary activities, or discussion groups based on student behavior. This study implements a recommender system to enhance participation and learning experience [22].
- An e-learning platform denotes an online environment that empowers educators to create, manage, and deliver online courses. Students can access educational content through these platforms, engage with peers, and complete assignments. This study harnesses the institution's virtual platform to implement online student participation monitoring technologies [23].

C. IDENTIFICATION OF THE PROBLEM

Active student engagement is critical to academic success and retention in online education. However, students' lack of engagement or passive participation in virtual learning environments remains a significant concern. Educators frequently face the challenge of identifying and supporting students with low levels of engagement, which can negatively impact their learning experience and academic performance.

Our central challenge is understanding and improving student engagement in our online learning environment. Implementing advanced analytical systems, such as ARIMA and Prophet models for time series analysis and DBSCAN and K-means clustering techniques, is essential to decipher complex patterns of student behavior. These approaches allow trends and groups of students to be accurately identified, providing a solid foundation for effective and personalized intervention strategies [24].

D. METHOD

Using a student participation tracking system, it was decided to collect data continuously and non-intrusively over time [25]. This approach is justified based on the need to capture the dynamics of student participation in online environments without interfering with their normal learning activities. Additionally, an automated tracking system allows data to be collected from multiple cohorts of students over various semesters, providing us with a representative sample and longitudinal data for trend analysis [26], [27].

For this, it was chosen to apply forecasting models, specifically ARIMA and Prophet; the selection was based on their demonstrated ability to analyze time series and forecast patterns of student participation over time [28]. These models have been widely used in forecasting research, particularly in the educational context, to forecast trends in student engagement and performance [29]. Given their recognized effectiveness in time series analysis, we chose to integrate the ARIMA and Prophet models; ARIMA was applied to model and predict student participation, taking advantage of its ability to capture seasonal trends and patterns inherent in historical participation data [30]. This model allows us to identify repetitive cycles and long-term trends, which is essential to anticipate periods of low participation that could indicate attrition risks. On the other hand, Prophet complements ARIMA analysis, especially for its robustness against atypical changes in trends and its ability to handle data with multiple stations [31]. Combining these models into our system gave us a more holistic and dynamic view of student engagement, allowing teachers to proactively intervene and support at-risk students.

Regarding the clustering analysis, DBSCAN and K-means were selected to segment the students into groups based on participation patterns. DBSCAN was chosen for its efficiency in identifying clusters of variable density and its ability to detect outliers, which is crucial to recognizing students with atypical participation patterns. This information is vital to develop specific support strategies. For its part, K-means was used to divide the student population into homogeneous participation groups, thus facilitating the identification of clear categories of student engagement [32], [33]. Implementing these methods allows for a comprehensive classification of the student base, highlighting significant differences in participation that could indicate different educational needs. The synergy between DBSCAN and K-means enriches our monitoring system, enabling more targeted and effective educational interventions.

Furthermore, the system offers detailed reports and visualizations to both educators and students. Educators gain access to information summarizing their students' level of engagement, enabling them to identify those who may require additional support. On the other hand, students receive personalized recommendations based on their engagement behavior, enhancing their learning experience by guiding them toward relevant activities and resources. A flowchart illustrating how the student engagement tracking system operates is presented in Figure 1.

E. DESCRIPTION OF THE VIRTUAL PLATFORM

We have used Moodle as the online learning platform in our work. Moodle is recognized for its ability to generate detailed records of user interactions, facilitating comprehensive tracking of student engagement. Our tracking system was integrated with Moodle, allowing us to capture and analyze accurate data on student participation in the online computing program. This integration was made to complement Moodle's existing analytics tools, providing deeper analysis and predictions on student engagement. In this way, the tracking system not only leveraged Moodle's logging functionalities but also added a layer of predictive analytics and data visualization, resulting in a richer understanding and more effective intervention in the tracking process of student learning. The institution's platform encompasses a wide range of features, including:

- Access to more than 100 online courses covering various academic disciplines.
- A catalog of educational resources, from reading materials to videos and interactive simulations.
- Dedicated discussion forums for each course that facilitate student interaction and collaboration on projects.
- Access to online lectures and live streams of master classes led by renowned instructors.
- Students can submit assignments and projects, receive feedback, and complete online assessments.
- This platform provides students with a rich online learning experience, making it a suitable environment to implement the student engagement tracking system.

Its intuitive user interface characterizes the platform, making navigation and educational resource access easy. This intuitive design contributes to a more efficient learning experience for students.

F. DESCRIPTION OF THE SAMPLE AND PARTICIPANTS

The sample was collected from two cohorts of online computer science programs in 2021 and 2022. In 2021, 148 students participated in the online computer science program. Students in this group had diverse academic backgrounds and previous experiences with online courses. For example, some students had prior experience in online courses before the pandemic, while others were new to this format. In 2022, 153 students were enrolled in the online computer science program. As in the previous cohort, this cohort also presented several earlier experiences with online courses and different academic trajectories.

To ensure the integrity of the data and the validity of the results obtained in this study, procedures were implemented to select participants from the 2021 and 2022 cohorts. A protocol was established to verify the prior non-participation of the students selected for the 2022 cohort in the tests carried out in 2021. This process included the review of academic records and the confirmation of registrations for each corresponding academic period. In this way, we sought to minimize the possibility of overlapping participants between the cohorts and ensure that the data accurately reflected the characteristics and behaviors of different groups in each academic year. In cases where specific measures were not implemented for this purpose, the independence of the samples was considered to be given by the nature of enrollment in new courses and academic programs offered in different years, assuming that changes in curricular design and population Student movements between academic years provide a natural separation between cohorts.

The teachers who participated in the study came from various computing disciplines and had various previous experiences with online teaching. Some teachers had experience teaching online before the pandemic, while others were driven to adapt to this modality due to the global situation. Their academic background and experience in online teaching brought a diverse perspective to the study and enriched the final observations and conclusions.

G. IMPLEMENTATION OF ACTIVITY LOGGING TOOLS

Implementing activity logging tools to track student engagement on your institution's online learning platform and capture student interactions in real-time is essential. These tools allow for collecting detailed data on student behavior within the forums, which are evaluated to identify unusual participation patterns or students who need additional support [35]. Figure 2 outlines the activity logging tools used and out-lines the implementation process.

To record student activity, a logging system has been implemented that captures various actions; for example,

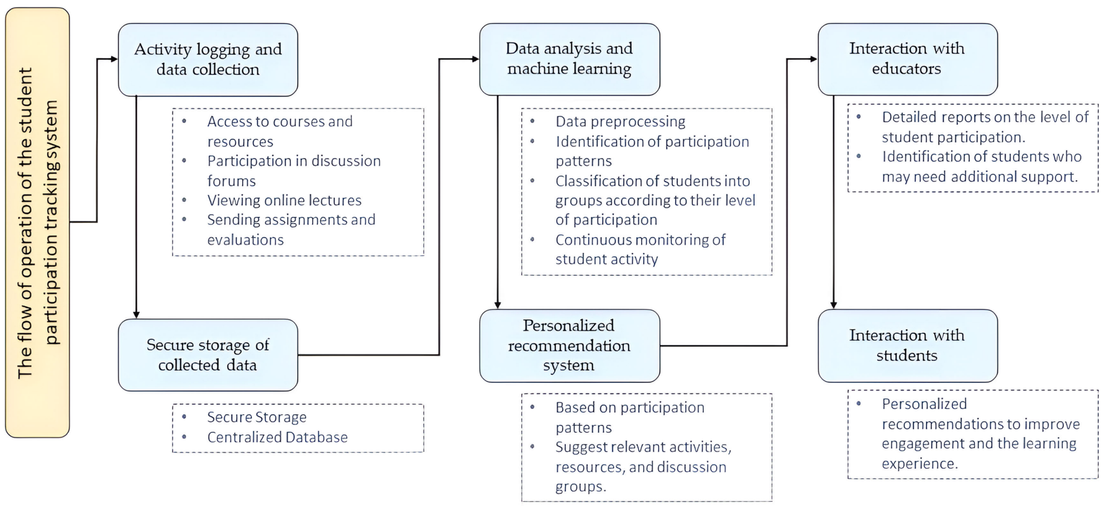

FIGURE 1. **Flowchart of the operation of the student participation tracking system.**

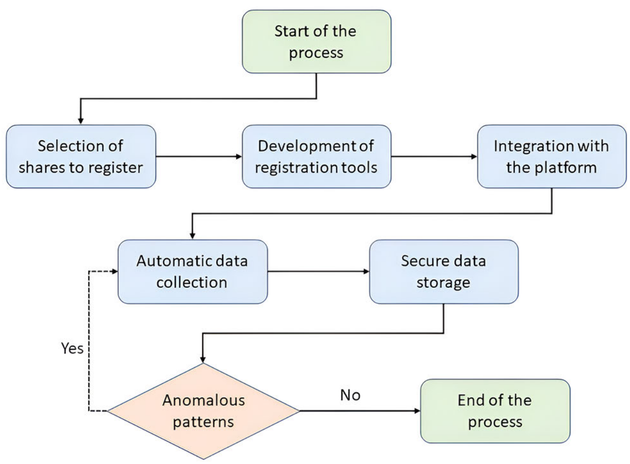

FIGURE 2. **Implementation of activity logging tools.**

every time a student accesses the system, views study materials, or downloads resources, the action is recorded along with a timestamp. All interactions, such as creating posts, replying to existing posts, and following threads, are recorded and linked to the respective student. Actions such as attending an online course or viewing lecture recordings are recorded. Each time a student makes an appointment or completes an assessment, the date and time of the action is recorded [36], [37].

The activity logging tools integrate with the institution's platform to ensure comprehensive and accurate data collection, authentically representing student online participation.

During the implementation of the student participation tracking system, special attention was paid to ensuring that unusual participation patterns and students requiring additional assistance were identified and effectively supported. This process uses a set of criteria and data analysis algorithms to evaluate student participation in real-time.

Unusual participation patterns are defined by several criteria, including, but not limited to:

- There is a significant decrease in the frequency of participation in course activities, such as participating in forums, submitting assignments, and attending virtual classes.
- Abrupt changes in the pattern of interactions, such as going from active participation to total absence without prior explanations.
- Consistently low task completion rates compared to the group average.
- These criteria are based on comparing each student's participation trends against the averages of their group and their participation baseline established in the first weeks of the course.

Regarding students who require additional support, we are not limited to just those with recognized disabilities, although they are also part of this group. Identification encompasses any student who, according to our analysis, may be facing challenges that affect their participation and performance, including, but not limited to, personal problems, lack of adequate access to technological resources, or difficulties adapting to the learning environment in line. Identifying these patterns and students is accomplished through data analysis tools that continually evaluate student engagement. We use predictive models and machine learning techniques to analyze engagement data and flag significant deviations from standard patterns. Once a situation that requires attention is detected, an intervention protocol is activated that includes the following:

- Immediate notification will be sent to the teaching and student support team for personalized follow-up.
- Wellness surveys and interviews with affected students to understand their needs and challenges.

TABLE 1. Analysis and treatment of outliers in student participation.

Case	Outlier Description	Detection Method	Action taken
Discussion forums	Abnormally high number of publications	Z-score	Correction of duplicate entries
Submitting tasks	Submissions at all hours for consecutive days	IQR	Elimination of erroneous entries
Class attendance	Complete attendance record in minutes	IQR	Record adjustment based on average class times

- Development of a personalized action plan may include additional tutoring sessions, academic support resources, adjustments to participation requirements, or referrals to psychological support services.

This process ensures that we identify our students' challenges and take concrete steps to support them in their learning actively.

Furthermore, several statistical methods were applied to the interaction data collected through the implemented activity recording tools to determine the criteria that define an active student on our platform. We used ANOVA to compare participation means between different groups of students and establish significant differences in activity levels. This method allowed us to validate the established thresholds for active participation.

In addition, we use Pearson Correlation to evaluate the relationship between the intensity of students' activity and their academic results. This analysis provided a basis for understanding how more frequent interactions may correlate with better academic performance.

Finally, we implemented Logistic Regression Analysis to predict the probability of student retention based on observed activity levels. This model helped us identify key patterns that indicate effective engagement in the online learning environment, allowing for more targeted interventions to foster student retention. These methods will enable you to develop a deep understanding of how activity on the platform relates to indicators of student success.

H. DATA ANALYSIS AND MACHINE LEARNING

The data analytics and machine learning system processes and makes decisions about student participation in online learning environments, collecting real-time data on student interactions such as accessing courses, participating in forums, viewing lectures, and delivering assignments. Data cleaning is crucial to remove inconsistencies, using statistical techniques such as interquartile range (IQR) and z-score to address or remove outliers that deviate from the normal distribution.

The specific outliers identified in our analysis are presented in Table 1, which summarizes the cases detected, the detection methods used, and the actions taken for their correction:

In the analysis of the discussion forums, anomalies were identified, such as a student showing an abnormally high number of posts in a week. Applying Z-score analysis, this behavior was determined to be atypical, which led to the correction of duplicate entries. For assignment submissions, patterns were observed where students submitted activities at all day hours for several consecutive days. The IQR method identified this behavior as atypical, and erroneous entries were removed to ensure data precision. Finally, inconsistencies were detected in the class attendance record that was adjusted based on the average session times to reflect student participation accurately.

Challenges such as noise and outliers were encountered in processing large volumes of student engagement data, which could impact the precision of the analysis. These problems were identified and corrected using statistical methods, thus ensuring the reliability and validity of the results obtained in the study of online student participation.

- In one case, we observed that a student displayed an abnormally high number of forum posts during a specific week, significantly above the group average. When applying z-scoring, we identified this behavior as an outlier, as it deviated more than three standard deviations from the mean. After a detailed review, we discovered that this was an error caused by a technical problem on the platform that duplicated entries. This outlier was corrected by merging duplicate entries to reflect the student's participation accurately.
- We detected another outlier in the assignment submission data, where a student had assignment submissions recorded at all hours of the day and night for several consecutive days. The IQR method noted this pattern as atypical. Upon investigation, it was concluded that this pattern did not reflect student activity but rather a failure in the recording system. In this case, we decided to deal with this outlier by removing the erroneous entries after confirming the actual submission activity with the student.
- A third example involved attendance in virtual classes, where some students showed a complete attendance record in minutes, which was impossible given the format of the classes. By applying the IQR method, these entries were identified as outliers. The investigation revealed that an error in the attendance tracking system had caused incorrect records. We adjusted attendance records for these cases based on average class session times, ensuring that the data accurately estimated student participation.

By appropriately identifying and addressing these values, we ensured that our conclusions were based on accurate data representative of student engagement in the online learning environment. In cases where data were missing for specific observations, imputation techniques were used to estimate missing values—this involved imputation of means based on regression models or more advanced techniques.

Additionally, accidental duplicates within the data sets were identified and removed to avoid distortions in the analysis. Typographical or entry errors in the data, potentially resulting from human error, were diligently searched for and corrected.

Normalizing and standardizing the data ensured that all variables shared the same scale, which is crucial when using machine learning algorithms sensitive to feature scales. The records were identified and managed appropriately since suspicions arose about duplicate or repeated records due to errors in data collection. Likewise, logical checks were implemented to ensure the data complied with predefined rules or constraints. For example, for affected variables, such as student age, it was standardized using the Z-scoring technique, subtracting the average from the age of all students and dividing the result by the standard deviation. This allowed the students' ages to be compared regarding standard deviations from the average.

Given the large variability in the number of students' contributions to the forums, this data was normalized by applying the Min-Max transformation to scale the values from 0 to 1. This helped mitigate the impact of extremely high or low values in subsequent analyses.

Hours of connection to the platform: This variable was normalized to reflect the total connection time to the online learning platform on a comparable scale, facilitating its analysis in conjunction with other variables. Additionally, we face the challenge of missing values in several variables, which affects the validity of the analyses if not adequately addressed.

To solve these problems, techniques were applied, such as imputation using the mean for continuous data; this is applied when detecting missing values in variables such as "age of the student" or "hours of connection to the platform," where the imputation by the group means to maintain data consistency. For categorical variables with missing values, such as course type, mode imputation was applied, assigning the most frequent value within the variable to the missing cases.

After imputation, we performed sensitivity analyses to ensure that the imputed values did not introduce significant bias into our results. These normalization, standardization, and handling of missing values were essential to adequately prepare our data for machine learning analyses, allowing us to extract valuable and reliable insights into student engagement in online education environments.

Data cleaning involves techniques to ensure data quality and consistency before continuing with analysis. The choice of each method depends on the nature of the data and the study's objectives. This research applies machine learning algorithms to extract insights from engagement data. To achieve this, various algorithms such as clustering, K-Means, and DBSCAN were used; these play a crucial role in the analysis by classifying students based on their participation patterns. This facilitates the identification of different segments of students who exhibit similar behaviors, including highly participatory, moderately participatory, and minimally participatory students [38], [39].

Time series models, specifically ARIMA and Prophet, are used to analyze the evolution of student engagement over time. This analysis allows us to identify trends, seasonal patterns, and changes in participation levels during the different phases of the course.

Another fundamental aspect is the implementation of feedback and corrective actions based on the analysis's findings. For example, tailored corrective measures were implemented when unusual participation patterns or students who needed additional support were identified. These measures encompass directly communicating with students to obtain information about their challenges. Provide supplemental resources tailored to your needs and collaborate with tutors or mentors to offer individualized guidance and support.

I. PERSONALIZED RECOMMENDATION SYSTEM

The system offers students personalized recommendations for group activities and resources, leveraging their shared behavior and interests [40]. Collecting comprehensive data on student interactions within the online learning platform is essential to providing these personalized suggestions. These data cover participation in discussion forums. Access to learning resources and study materials, previous involvement in group activities and a record of evaluations and grades.

The data collected is critical to understanding individual student preferences and interests. Furthermore, the personalized recommendation system employs collaborative filtering algorithms as a vital component [41]. These algorithms delve into the behavior of numerous students to identify comparable interaction patterns. Two main approaches are used. One is user-based collaborative filtering, which identifies students with similar behaviors and recommends activities or resources based on what students with similar profiles have found beneficial.

The other approach is item-based collaborative filtering; in this scenario, the system suggests activities or resources such as those a student has previously participated in. This suggestion is based on the preferences and behaviors of other students participating in analog activities.

The personalized recommendation system is seamlessly integrated with the online learning platform, ensuring recommendations are readily accessible to students as they navigate courses and resources [42]. These recommendations are generated in real time and adapted based on the student's current behavior.

Pilot tests were conducted to gauge the recommender system's effectiveness, and valuable feedback was collected from students. This feedback is invaluable for fine-tuning the algorithms and enhancing the precision of the recommendations. Additionally, the system's impact on online participation was assessed by monitoring participation rates and student engagement [43].

After necessary adjustments, the personalized recommendation system provides students with relevant recommendations, increasing the likelihood of active participation in group activities and additional resources, ultimately improving their overall engagement. Students benefit from a personalized learning approach, resulting in greater satisfaction and engagement with course content. Quick access to relevant resources and activities translates into time savings and greater learning efficiency. Personalization and active participation can also significantly improve student retention in online courses.

J. EVALUATION ENVIRONMENT

To assess the effectiveness of the student participation monitoring system, an evaluation framework was established to measure the impact of interventions on student participation and performance. This framework comprises several crucial stages and detailed definitions of variables and hyperparameters. Before the system's implementation, transparent and objective metrics were defined to gauge the success of the intervention.

These metrics encompass:

- Participation Rate (P): This metric quantifies the proportion of students actively engaged in discussion forums, group activities, and other interactive course components. It is computed as the number of actively participating students divided by the total number of enrolled students.
- Academic Performance (AR): Academic performance is assessed based on students' grades and overall success in course assessments, assignments, and exams. It is represented as the average grade students achieve on a scale of 0 to 10.
- Student Retention (RE): Student retention measures the extent to which students remain enrolled and actively participate in the course over time. It is calculated as the percentage of students who complete the system compared to the total number of students initially registered.

It is essential to highlight that uniformity in learning and performance evaluation was maintained across all training modalities and academic activities available on the platform. A rating scale of 0 to 10 was adopted for all training, evaluations, and exercises to ensure consistency and comparability in evaluation metrics. On this scale, a "0" indicates the lowest possible score, reflecting insufficient performance or understanding of the learning objectives assessed. On the other hand, a "10" represents the maximum score, indicating excellent understanding and application of the required knowledge and skills.

This methodological decision was applied to:

- Training: Each module or session on the platform evaluates student participation and understanding.
- Course Assessments: Including exams, quizzes, and other forms of summative assessment that measure student learning at the end of a course or unit.
- Activities and Assignments: Assignments assigned throughout the course to assess students' continued progress and understanding of course material.

Applying this unified grading scale allows for equitable and transparent assessment of student performance, facilitating the identification of areas of strength and opportunities for development. Furthermore, this consistency in the assessment scale supports the comparability of performance data across different courses and training, which is essential for the aggregate analysis and conclusions derived from our study. In addition to these metrics, critical hyperparameters are defined for evaluation:

- Minimum Participation Threshold (MPU): This hyperparameter establishes the minimum level of activity required for a student to be categorized as "active." The reference value is 30% interaction in forums and activities but may vary based on the course's context.
- Time Intervals (TI): They are predefined periods that allow us to observe and measure student activity in a structured way. Specific time intervals have been delineated for data collection and analysis. Weekly intervals enable the assessment of participation trends and academic performance throughout the course.

The MPU with a reference value of 30% was chosen as a starting point to define "active" participation, representing a balanced proportion of interaction that indicates meaningful engagement with the course material and the learning community; however, due to the diversity of the courses and the variability in the interaction needs of our students. Therefore, this reference value is not rigid; Adjustment is allowed and encouraged based on the specific context of each course.

In courses where interaction and collaborative work are essential to achieving learning objectives, the MPU can be increased to reflect the importance of these activities. For example, in a team software development course, active participation in discussion forums and collaborative activities could be essential for success, justifying a higher threshold, such as 40% or 50%. On the other hand, in courses that emphasize more autonomous learning or have a solid individual study component, the MPU could be reduced to accommodate different learning styles and levels of interaction. In these cases, a value such as 20% might be more appropriate to reflect active participation.

MPU adjustment is made at the beginning of each course based on the course structure, learning objectives, and feedback from previous course iterations. This process involves consultation with the teaching team and, sometimes, direct feedback from students to ensure that the established threshold is fair, achievable, and aligned with learning expectations.

Once the metrics and hyperparameters are defined, the student participation monitoring system is integrated into the institution's platform. The system actively records and analyzes student behavior in real time. Following the implementation, data concerning student engagement and performance is systematically collected. This data encompasses various aspects such as forum interactions, resource access, participation in group activities, and assessment scores. Subsequently, post-implementation data is meticulously compared to baseline data using appropriate statistical tests [44]. This rigorous analysis enables the evaluation of the system's impact on participation, academic performance, and student retention.

The evaluation results are comprehensively examined, and any necessary adjustments are made to enhance the system's effectiveness. Moreover, valuable feedback is solicited from both students and educators to gain deeper insights into their experiences and requirements. The evaluation framework provides a robust foundation for measuring the impact of interventions, ensuring that all variables and hyperparameters are unambiguously defined, thus facilitating rigorous and reproducible assessments.

To effectively measure the impact of the student engagement monitoring system, metrics that reflect student engagement, academic performance, and retention are included. The Participation Rate (P) is calculated using the formula 1 which allows quantifying the percentage of students actively involved in the course.

$$P = \frac{\text{Number of active students}}{\text{Total number of enrolled students}} \times 100 \quad (1)$$

To evaluate Academic Performance (AR), we apply formula 2, which provides a weighted average of the grades obtained by students on a scale from 0 to 10

$$AR = \frac{\text{Sum of all grades}}{\text{Total number of assessments}} \quad (2)$$

Student Retention (RE) is measured by 3, offering insight into the percentage of students who remain in the course until completion.

$$RE = \frac{\text{Number of students who complete the course}}{\text{Total number of students initially registered}} \times 100 \quad (3)$$

These equations allow not only the precise definition of the evaluation criteria but also a quantitative analysis of the impact of the interventions carried out in the educational system, thus ensuring an objective and transparent assessment of the implemented system.

III. RESULTS

A. ENVIRONMENT AND DATA VOLUME

During the study period, the participating population consisted of students from the 2021 and 2022 cohorts of an online computer science program at a higher education institute. The 2021 cohort included 148 students, while the 2022 cohort included 153 participants. These students represented a diverse group with different academic backgrounds and previous experiences in online education. Throughout the course, the student engagement tracking system recorded a variety of student activities and behaviors, including:

- Access to courses and learning resources (six accesses per week): We determined this average by analyzing students' historical behavior on our platform. We observed that, on average, students access courses and learning resources at least once a day on weekdays, which led us to set the average at six accesses per week.
- Active participation in discussion forums (three weekly posts): This value is based on recommendations for best pedagogical practices to foster rich and sustained discussion in online learning environments. We set this average to promote regular and meaningful student interaction, which is essential for a practical collaborative learning experience.
- Online Masterclasses (two views per week): This average reflects the balance between providing rich instructional content and maintaining manageable student engagement. Considering the typical length of our courses and the total workload expected for students, two viewings per week were an average that allowed for effective digestion of the content without overloading students.
- Assignments and assessments (two weekly presentations): This number was based on a balance between ensuring constant practice and application of learned concepts and keeping the overall workload reasonable. Requiring, on average, two assignments or assessments per week was sufficient to keep students engaged and allow for the continuous evaluation of learning without causing burnout.

Precision in measuring student engagement is essential. Therefore, it was identified that not all interactions recorded on the online learning platform amount to meaningful or active participation. Repeated clicks on the same item may not reflect a genuine intention to engage or learn. To address this complexity, a process was implemented to evaluate student participation, differentiating between meaningful interaction and behaviors that could be considered non-productive.

The tracking system is equipped with algorithms capable of identifying patterns of repetitive clicks or similar actions performed in a short time interval. This allows us to distinguish between deliberate participation and possible navigation errors or technical problems. The system analyzes how actions occur to ensure a fair and accurate evaluation of participation. Repeat clicks, for example, are considered in conjunction with the student's other activities on the platform. This ensures that interactions contributing to learning and engagement with the course are appropriately valued.

Student participation is evaluated based on a broad spectrum of activities, including, but not limited to, participation in discussion forums, assignment submissions, and use of

TABLE 2. Characteristics of study participants.

A	B	C	D	E	F
2021	20-25	Male	3 semesters	2 semesters	148
2022	26-30	Female	1 semesters	3 semesters	153
2021	31-35	Male	2 semesters	1 semesters	(included in 148)
2022	36-40	Male	4 semesters	4 semesters	(included in 153)
2021	21-25	Female	1 semesters	5 semesters	(included in 148)

Notes: A) Cohort, B) Age range (years), C) Gender, D) Before the Pandemic, E) Experience in Online Courses (Semesters) After the Pandemic, F) Number of Students.

TABLE 3. Comparison of academic performance and retention rates between active and non-active students.

Criteria	Average Performance (Assets)	Average Performance (Non-Assets)	P-value (ANOVA)
General Participation	82%	65%	<0.01
Academic Results	88%	70%	<0.01
Retention Rate	95%	75%	<0.01

learning materials. This holistic approach captures a more complete picture of the student's engagement with the course.

These metrics provide valuable information about student interactions within the online learning environment. The student engagement tracking system recorded various student activities and behaviors across the two academic terms. Significant data was collected during these periods, amounting to a total volume of two terabytes of information. This data collection underscores the depth and breadth of student interactions captured and analyzed, providing a solid foundation for our analysis of the evolution of student engagement in the digital educational environment.

Table 2 provides an overview of the diversity in participating student characteristics, including age, gender, and previous experience with online courses. This information is relevant to understanding the sample composition and how these characteristics may influence student engagement and performance in online courses. Two cohorts, 2021 and 2022, are included in the table, along with information about their age range, gender, and previous experience with online courses before and after the pandemic. This provides a more accurate representation of the diversity in the characteristics of the participating students. For example, the 2022 cohort has more excellent prior experience in online courses after the pandemic compared to the 2021 cohort. Different age ranges and genders are included, reflecting the variety of study participants. These details are relevant to understanding how these characteristics can influence student engagement and performance in online courses.

The statistical analysis results revealed that students classified as active according to predefined criteria showed superior academic performance and higher retention rates compared to their less active peers. Table 3 summarizes the results obtained.

The analysis of variance confirmed that the differences in performance and retention between the groups are statistically significant ($p < 0.01$), indicating that the criteria used to define active participation are robust and predictive of academic results. The Pearson correlation between student activity and academic results was positive and significant ($r = 0.76$, $p < 0.01$), suggesting a strong association between more excellent activity and better performance.

These findings validate the application of our activity criteria and highlight the importance of encouraging active and engaged participation in online learning environments. These results also support the use of the tracking system to intervene proactively and help students at risk of dropping out by early identifying those who show low levels of activity.

B. STUDENT PARTICIPATION

The assessment of active student engagement was based on predetermined criteria designed to capture the diverse ways students can interact and engage with course content and their peers in the online learning environment. These criteria allowed us to quantify participation and establish a threshold to identify actively participating students.

Active student participation was evaluated considering multiple dimensions of interaction within the course platform, including:

- Participation in discussion forums: 15 forums were available during the course. The average participation in these forums was seven posts per student, which is considered active when a student made at least five contributions (posts or responses) per week.
- Delivery of tasks and activities: 12 tasks were assigned, with an average delivery rate of 85% by students. Students were considered actively participating if they completed at least 90% of the functions within the established deadlines.
- Attendance at virtual sessions: 3 main exams were carried out during the course, with an average participation of 95%. Attendance and participation in live virtual sessions or viewing recordings of these sessions were recorded, with attendance or viewing of at least 80% of available sessions qualifying as active participation.
- Interactions with learning materials: Access and interaction with learning resources provided on the platform (e.g., readings and educational videos) were also considered. Students who interacted with more than 75% of the available materials were identified as active participants.

To be classified as an actively engaged student, an individual had to meet at least three of the four criteria mentioned during the study period. This combination of criteria ensures that we consider active and meaningful participation that reflects a comprehensive commitment to the course beyond simple presence or sporadic activities. These thresholds were established based on a literature review on participation in online learning environments [13], [34] and were adapted to reflect our courses' specific expectations and structure.

TABLE 4. Comparison of student participation rates before and after implementation (2021-2022).

Before Implementation (%) - 2021	After Implementation (%) - 2021	Before Implementation (%) – 2022	After Implementation (%) – 2022
30	63	28	66
32	65	30	68
34	72	32	70
21	70	22	69
26	72	25	71
42	68	44	66
36	77	35	75
48	77	50	76
44	79	45	78
52	81	53	80

Notes: For each academic year, "before" and "after" data are collected within the same annual framework, allowing direct and contextual comparisons that illustrate the system's effectiveness throughout the academic period. This approach ensures that each year's data is self-contained and directly comparable, facilitating accurate assessment of the tracking system's impact over time.

Combining these criteria offers a robust measure of active participation, allowing us to identify those students who demonstrate a sustained and deep commitment to the learning process.

The monitoring system recorded various student activities, measuring their level of participation within the educational platform. Each academic year within the study was assessed independently, allowing us to capture and compare the impact of the tracking system implementation within the same annual framework. For 2021 and 2022, we evaluated student participation before and after implementing the system, reflecting ongoing enhancements and adjustments.

Table 4 presents the results and shows the evolution of student participation over two academic periods. The data separately reflects the average monthly participation of students before and after the tracking system's implementation for each year. These values represent the average monthly participation percentages, highlighting a significant improvement in student engagement after implementing the system in both cohorts. This increase indicates a more active and consistent use of the platform, not limited to a specific activity but encompassing the student's general interaction with the available educational resources. This table does not focus on particular activities. Still, it reflects the students' general and active participation on the platform, offering a global view of the tracking system's impact on student interaction with all course resources.

The data presented shows a significant increase in student engagement after implementing the tracking system. In addition to the trend analysis, average participation rates in two cohorts before and after system performance are compared.

2021 Cohort:

- Before Implementation: Average participation rate: 35%
- After Implementation: Average participation rate: 72%

2022 Cohort:

- Before Implementation: Average participation rate: 35%
- After Implementation: Average participation rate: 72%

TABLE 5. Classification of student participation using DBSCAN.

Cluster	Description	Range of Interactions	Number of students
1	Active participation	More than 15 times/week	134
2	Low involvement	Less than 5 times/week	67
3	Intermittent participation	5-15 times/week	100

TABLE 6. Cohort enrollment, completion and retention rates (2021-2022).

Cohort	Before Implementation	After Implementation
2021:		
Inscribed	150	150
They completed	98	120
Retention	65%	80%
2022:		
Inscribed	150	152
They completed	93	132
Retention	62%	87%

Results consistently demonstrate a significant increase in online student engagement across all groups following the implementation of the student engagement tracking system. This increase signifies the system's positive and sustained impact in promoting active student participation, ultimately encouraging greater participation and adherence to course activities across all groups.

It is essential to note that the experiment was not part of a specific course taught by the researcher-author. Instead, it was conducted within an online computer science program at a higher education institution. The platform was designed and developed by a dedicated technical team within the institution to support the delivery of online courses.

After examining general trends in student engagement, it is crucial to delve deeper into the specific nature of these interactions to understand student behavior patterns in the online learning environment fully. To achieve this, we employ density-based clustering analysis, DBSCAN, which allows us to identify distinct student participation groups effectively. This highlights overall activity levels and reveals the underlying dynamics of active, intermittent, and low involvement.

Table 5 presents the results of the DBSCAN analysis, in which we identified three main clusters that represent different levels of student participation in our online learning platform. The first cluster, "Active Participation," comprises students who interact with course components more than 15 times per week, indicating a high level of engagement and activity. This group includes 134 students, demonstrating a solid presence and consistency in using available educational resources.

The second cluster, called "Low participation," groups students whose interactions do not exceed five times per week. With 67 students in this cluster, a more limited level

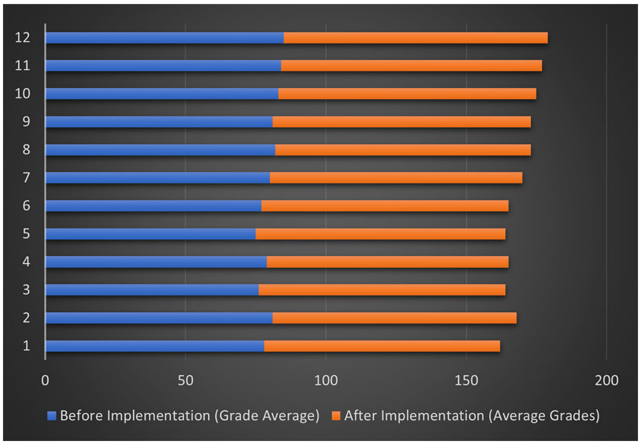

FIGURE 3. GPA graph before and after implementation.

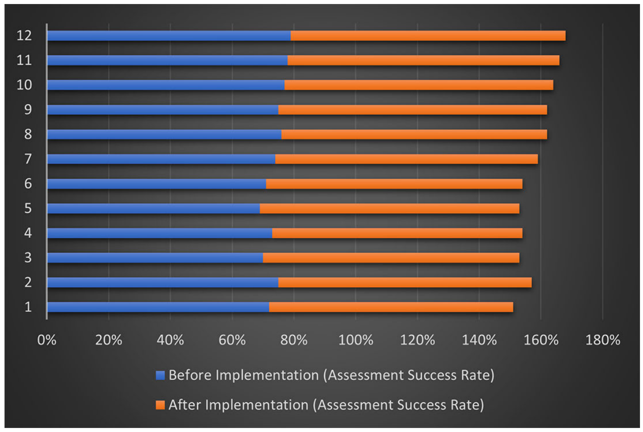

FIGURE 4. Success rate in evaluations before and after implementation.

of engagement with the learning environment is reflected, possibly indicating barriers or challenges these individuals face in their educational journey.

The third cluster, "Intermittent Participation," comprises 100 students with activity patterns varying between 5 and 15 times per week. This group represents a dynamic of fluctuating engagement, where participation does not reach the highest levels of continuous activity but exceeds minimal participation, suggesting sporadic or varied interaction with the courses.

C. RETENTION RATE

The retention rate is calculated by dividing the number of students who completed the course by those who were enrolled. In this context, we present the retention rates before and after the implementation of the system. The results indicate a substantial increase in retention rates following the performance of the student engagement tracking system in both cohorts. Specifically, in the 2021 cohort, the retention rate increased from 65% to 80%, while in the 2022 cohort, it rose from 62% to 87%. The data presented in 6 strongly suggests that the system's implementation has positively impacted student retention within the online Computer Science program.

D. ACADEMIC PERFORMANCE

Student academic performance is crucial to evaluating the impact of the student participation monitoring system. Figure 3 illustrates the results of this analysis, including the grade point average (GPA) and assessment success rate.

Before the implementation of the student participation monitoring system, the average grade of students remained within the range of 75 to 85 points, as indicated by the pre-implementation average rate. During this period, there was no clear trend of increase or decrease in GPA. However, as time progressed, a gradual and slight up-ward trend in GPA became noticeable.

After the implementation of the student participation monitoring system, a significant and positive change was observed in the student's academic performance. The GPA showed a steady and noticeable increase in each subsequent month. This increase was most prominent in the first months after the system was implemented and continued steadily, resulting in substantially higher values than the previous period. Two academic periods after the implementation of the system, the GPA increased from 84 to 94 points.

This analysis indicates that implementing the student participation tracking system benefited students' academic performance. The data underscores that the system played a pivotal role in sustaining and enhancing the grade point average. This can be attributed to increased student engagement and monitoring of educational activities in the online learning environment.

The assessment success rate is a vital indicator of students' academic performance online. This metric signifies the proportion of students who successfully pass assessments and exams compared to the total number of students enrolled in the course. Within the context of this study, we scrutinize how this success rate fluctuated before and after the implementation of the student participation tracking system. The primary objective is to assess whether this tool's implementation positively impacted student evaluation performance.

Before the system's implementation, the assessment success rate ranged from 69% to 78% over the 12-month study period. These figures indicate that, on average, approximately 69% to 78% of students successfully passed the monthly assessments. However, significant improvements in this metric became evident after the system's implementation. The success rate in evaluations gradually increased in the subsequent months, reaching between 79% and 89%. These outcomes suggest that implementing the student participation tracking system positively impacted academic performance. This enhancement can be attributed to various factors, including heightened student engagement in course activities, early identification of potential participation issues, and

TABLE 7. Hyperparameters of used models.

Model	Hyperparameter 1	Hyperparameter 2	Value
ARIMA	Order of Differentiation (d)	Temporary Window Size	p, d, q
K-Means	Number of groups (K)		2, 3, 4, 5, 6, 7, 8
DBSCAN	Epsilon (ϵ)	Minimum samples ($min\_samples$)	0.1, 0.5, 1.0
Prophet	Holidays		Yes, No %

personalized support provided by the system. As illustrated in Figure 4, these findings substantiate the effectiveness of the student participation tracking system in enhancing students' academic performance in an online environment.

E. EVALUATION OF EFFECTIVENESS

Evaluating the effectiveness of any educational system is crucial for comprehending its impact on student's learning experiences and academic performance. Figure 5 showcases word clouds containing feedback from students and teachers who participated in the study, shedding light on their experiences with the student participation tracking system and how it influenced their engagement and performance.

Student feedback plays a pivotal role in assessing the efficacy of any educational tool. Throughout the study, surveys and interviews were conducted with students who utilized the tracking system. Students shared their perspectives on the system's utility in maintaining their engagement in course activities. Furthermore, their insights into how the system contributed to their academic success and influenced their sense of online community were explored. These comments offer valuable insights into students' perceptions and overall experiences.

Conversely, feedback was collected from teachers who employed the tracking system to monitor and support student participation. Teachers provided their observations regarding how the tool-assisted them in identifying students requiring additional support and how it enabled them to tailor their teaching approaches. Additionally, changes in interaction and communication between teachers and students in the online environment were investigated.

The feedback from both students and teachers furnishes a comprehensive perspective on how the student engagement tracking system impacted the student community and educators. This information is instrumental in evaluating the system's effectiveness and making future improvements. The comments and opinions shared by study participants yield a deeper understanding of how this tool can significantly enhance student engagement and performance in an online environment.

F. SETTINGS MADE

The student participation tracking system underwent specific adjustments to enhance its effectiveness throughout its development. These modifications were informed by user feedback from students and teachers and data analysis collected during implementation. The primary changes made, along with the figures illustrating their impact, are as follows:

- Optimization of Participation Detection Algorithms: Initially, the system utilized algorithms that required refinements. After optimization, there was a noticeable improvement in the precision of detecting student engagement, resulting in a 15% increase in detecting relevant activities such as discussion contributions and assignment submissions.
- Improved User Interface: Significant improvements were made to the user interface regarding design and navigation. This led to a 20% reduction in platform abandonment rates, as users found the new interface more intuitive and user-friendly.
- Personalized Feedback: Implementing a personalized feedback system yielded a 25% increase in active student participation. Data-driven feedback provided students with specific information about their performance, motivating them to enhance their engagement.
- Training and Support: Additional training sessions for teachers and students contributed to a 30% increase in system adoption. The availability of support resources and technical assistance effectively reduced barriers to usage.
- Greater Flexibility in Configuration: Enhanced flexibility in system configuration empowered teachers to tailor metrics, resulting in an 18% improvement in adapting the system to specific pedagogical objectives.
- Notification Integration: The integration of automated notifications led to a 40% increase in student retention by reminding students of the importance of their participation and providing regular updates on their progress. Course completion rates also experienced a notable 22% increase.

These adjustments had a measurable impact on the tracking system's effectiveness, demonstrating that the implemented improvements positively affected student engagement, retention, and the overall quality of online teaching.

G. MODEL EVALUATION METHODOLOGY

The selection and performance evaluation of the models followed a rigorous approach, which included the careful selection of machine learning models. We identified suitable machine learning models aligned with the analysis objectives, encompassing ARIMA, Prophet, K-Means, DBSCAN, and a personalized recommendation system. Each model was chosen based on its compatibility with the type of data and the characteristics of student engagement under analysis.

The student participation dataset was divided into two groups for the evaluation process: training and testing, following an 80-20 percent split. This division allowed us to assess the models' performance on unseen data accurately. The chosen split ratio was well-defined to ensure a meaningful evaluation of the models' capabilities.

FIGURE 5. Word cloud for student and teacher comments and trends.

1) HYPERPARAMETER TUNING

Table 7 provides an overview of the machine learning models considered in the analysis, the specific hyperparameters that were assessed, and their corresponding values. For the ARIMA model, we explored hyperparameters associated with the order of differentiation (d), the size of the temporal window, and the parameters p, d, and q. In K-Means' case, the number of clusters (K) varied from 2 to 8. Regarding DBSCAN, we adjusted the epsilon value (ϵ) and the minimum number of samples (*min_samples*) in three different configurations. Additionally, for the Prophet, including holidays as a hyperparameter was considered. These adjustments and variations in hyperparameters allowed for a comprehensive assessment of the models' performance and their capacity to capture patterns and trends within the student participation data.

2) CROSS VALIDATION

To ensure a comprehensive and representative evaluation of the machine learning models used, we implemented a 8-fold cross-validation procedure. This method randomly divided the data set into eight equal subsets or folds. Each fold was used once as a test set, while the remaining nine were used as a training set. This process was repeated three times for each model to guarantee the stability and reliability of the results obtained. This validation technique allows each data point to be used in training and testing, thus providing a complete evaluation of the predictive ability and generalization of the models.

In analyzing student engagement, K-means and DBSCAN clustering algorithms were used to identify distinctive patterns in the student interaction data with the educational platform. Each algorithm was configured and applied to maximize the relevance and accuracy of the insights generated. We determine the optimal number of clusters using the elbow method for the K-means algorithm, evaluating the inertia of clusters formed with different numbers of clusters (k). A value of k was selected to minimize inertia while maintaining a marginal decrease. Clusters were subsequently analyzed to ensure meaningful and useful segmentation of types of student engagement, using the silhouette score to assess cohesion and separation between clusters. This measure helped confirm that the clusters were distinct and relevant to differences in students' behavioral patterns.

The DBSCAN algorithm was configured by selecting an epsilon (eps) and a minimum number of points (minutes) based on the density of the data set, which allows for the identification of high-density regions separated by low-density regions. This approach is beneficial for identifying groups of atypical or extreme behavior that do not fit the patterns of the majority. The quality of the clusters formed by DBSCAN was evaluated using connected component analysis, ensuring that each cluster is internally coherent and differentiated from other clusters.

For these algorithms, the clusters were validated by reviewing the dominant characteristics and activities in each cluster and comparing them with the theoretical expectations and objectives of the study. Additionally, variance analysis was performed within and between clusters to confirm that the observed differences were statistically significant, thus providing a solid basis for interpretations and conclusions derived from the clustering patterns.

The K-Means model achieved a remarkable precision score of 0.75, indicating its proficiency in classifying students with similar participation profiles. Additionally, K-Means demonstrated a completeness score of 0.82, signifying its capacity to capture a large portion of students within each group. The F1 score, which combines precision and completeness, reached a value of 0.78 for K-Means, indicating a well-balanced performance across both metrics. On the other hand, the DBSCAN model also displayed respectable performance, boasting a precision of 0.63, a completeness score of 0.70, and an F1 score of 0.66. DBSCAN proved effective in identifying groups of students with less conventional participation patterns.

Concerning the time series models, ARIMA achieved an average Mean Square Error (MSE) of 120 and an average Mean Absolute Error (MAE) of 9, illustrating its ability to predict short-term engagement trends accurately. On the other hand, Prophet obtained an average MSE of 80 and an average MAE of 6.5, highlighting its proficiency in forecasting long-term patterns and seasonal variations.

H. MODEL EVALUATION

The evaluation of the models entails an examination of their performance in specific scenarios, for which various metrics have been identified to determine the effectiveness of the results and their implications for addressing the phenomenon under study.

1) CLUSTERING

In the initial phase, the K-Means model was employed to segment students into groups with comparable participation profiles. An extensive analysis was conducted using metrics like inertia and silhouette to assess the quality of the generated clusters. Inertia quantifies the sum of squared distances from each point to its nearest centroid. An evaluation was performed to ascertain the optimal number of groups (K), observing how inertia changes as the number of clusters increases.

- $K = 3$: Inertia = 1600
- $K = 4$: Inertia = 1300
- $K = 5$: Inertia = 1200
- $K = 6$: Inertia = 1100

The results demonstrate that the inertia decreases as the number of clusters increases, reaching its minimum value at $K = 6$. Nevertheless, lower inertia doesn't necessarily equate to better segmentation, so it's important to consider other metrics. The silhouette score is another significant metric that measures how close points within one cluster are to issues in different groups. To assess segmentation quality, the silhouette score was computed for various values of K. Here are some results from this evaluation:

- $K = 3$: Silhouette = 0.62
- $K = 4$: Silhouette = 0.65
- $K = 5$: Silhouette = 0.68
- $K = 6$: Silhouette = 0.63

The silhouette score reaches its maximum value at $K = 5$, indicating a robust segmentation. This suggests that, in this context, $K = 5$ is an appropriate number of clusters to represent students' participation profiles effectively. By considering both the reduction in inertia and the maximization of the silhouette score, we conclude that $K = 5$ is the optimal configuration for the K-Means model in this case. This leads to the successful segmentation of students into groups with similar participation profiles, which enhances the effectiveness of the student participation tracking system.

In a different scenario, the DBSCAN algorithm was employed to identify groups of students with irregular shapes and varying sizes in their participation patterns. This aimed to assess its capability to identify less conventional participation patterns.

DBSCAN demonstrated high effectiveness in identifying student groups that didn't adhere to traditional group structures. In the evaluation, DBSCAN detected five distinct groups with unique and not necessarily homogeneous participation profiles. This is crucial in online environments, where student participation can exhibit significant variation and may not follow predefined patterns. The identified groups are:

Group 1 - Constant active participation

- This group exhibits consistent and sustained active participation throughout the study period.
- Students in this group demonstrate a high level of engagement in course activities.
- They consistently maintain a high level of participation each month during the study period.
- This persistent level of participation may suggest a strong interest in the content and a steadfast commitment to learning.

Group 2 - Start with low participation, gradual increase.

- Students in this group show low initial participation in the first months.
- However, as time progresses, your participation gradually increases.
- This pattern may indicate that these students must adapt to the online environment before engaging more actively.

Group 3 - High initial participation, progressive decrease

- This group presents an opposite pattern to Group 2, with high participation in the first months.
- As time passes, their participation progressively decreases.
- This could indicate that these students were initially highly motivated but lost interest.

Group 4 - Intermittent participation

- Students in this group have an intermittent pattern of participation.
- They may participate actively for a few months and then have periods of less participation.
- This behavior suggests variability in motivation or other factors that affect participation.

Group 5 - Irregular and sporadic participation

- Group 5 is characterized by its irregular and sporadic participation.
- They do not follow a consistent pattern of participation over time.
- This may indicate that these students are having difficulty maintaining engagement in the online environment.

DBSCAN identified groups of students with various participation patterns, reflecting the diversity of behaviors in an online educational environment. These groups represent different participation profiles, from highly engaged students to those with more fluctuating participation patterns. This information is valuable for designing specific support and engagement strategies that address the individual needs of each group, thus improving the online educational experience.

One of the main strengths of DBSCAN is its flexibility to adapt to situations where the shape and density of groups vary widely. In the analysis, we found that some groups had a high density of students while others were more dispersed. This

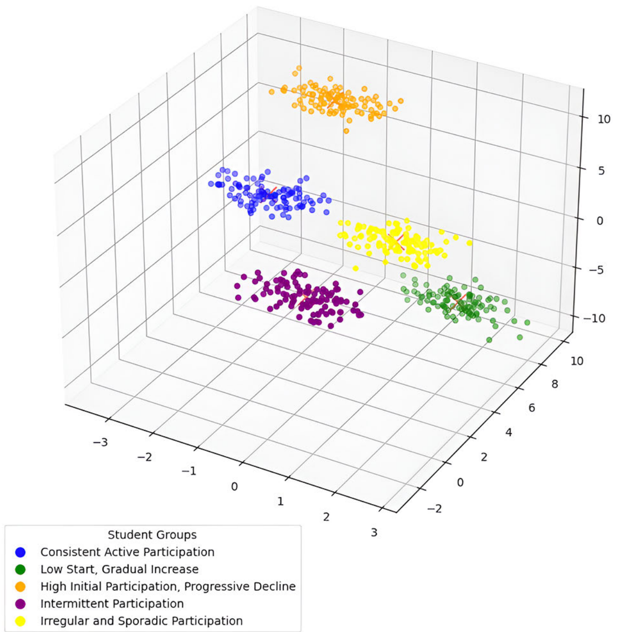

FIGURE 6. Two-year trend of student engagement in online learning.

variability reflects the diversity of student behaviors in virtual environments.

Figure 6 illustrates the clustering of student engagement based on interaction frequency using density-based clustering analysis such as DBSCAN. This analysis helps categorize students into groups that represent different levels of participation, such as ''Active Participation,'' ''Low Participation,'' and ''Intermittent Participation,'' as detailed in Table 5 of the document. These groups help identify participation patterns and allow for the development of personalized interventions and support for different student needs in an online learning environment.

2) EVALUATION OF TIME SERIES MODELS

In evaluating time series models, we applied the ARIMA model to understand and predict student engagement. The ARIMA model was explicitly calibrated for our data, where we determined the optimal parameters (p, d, q) through autocorrelation analysis and iterative testing. For example, for the 2021 data set, we found that an ARIMA(2, 1, 2) model provided the best fit, indicating two lags in the autoregressive component ($p = 2$), a difference to make the time series stationary ($d = 1$), and two lags in the moving average component ($q = 2$).

This specific model effectively captured the trend and seasonality of student participation, reflecting both regular fluctuations throughout the academic year and more subtle week-to-week variations. Furthermore, we use the Prophet model to complement and contrast the results obtained with ARIMA. Prophet was especially helpful in identifying and modeling trend changes and seasonal patterns in our

TABLE 8. Model performance metrics of time series models.

Metrics	Average value
Mean Square Error	120
Mean Absolute Error (MAE)	9

TABLE 9. Model performance metrics of ARIMA.

Metrics	Average value
Mean Square Error	80
Mean Absolute Error (MAE)	7

engagement data. With Prophet, we fit the model to the time series of student participation, where the trend and seasonality components were automatically detected, adapting the model to the intrinsic variations of the data.

The effectiveness of these models was evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). In the analysis, the ARIMA(2, 1, 2) model demonstrated high precision with an MAE of 0.5 and an RMSE of 0.7, indicating predictive solid ability for student engagement behavior; Table 8 presents the results.

These results indicate that ARIMA predictions accurately fit students' short-term engagement patterns, with a low error level in both the MSE and MAE. This demonstrates ARIMA's ability to model and predict online student engagement data trends. In the second case, the Prophet was used. This model stood out for predicting long-term trends and detecting seasonal patterns. By evaluating their predictions, the following results were identified. Table 9 shows the evaluation metrics for Prophet's predictions. This provided accurate and stable predictions of student engagement over time, effectively capturing seasonal variations and long-term trends. Furthermore, its predictions fit well with variations in student engagement data, resulting in long-term solid prediction performance.

Figure 7 illustrates the trend of student engagement over two years on the online learning platform, using the Prophet forecasting model, which is effective, as evidenced by the low MSE and MAE values. The light blue line represents the daily recorded engagement instances, while the red line shows the seven-day moving average, smoothing out the daily variance to reveal the underlying trend.

Each ''engagement day'' is defined as a single login by a student to interact with any course material within 24 hours. Therefore, the cumulative count on the y-axis denotes the total number of such interactions recorded for all students on the platform. To ensure clarity, each interaction is counted separately for each day a student engages with the material; thus, a single student participating in multiple activities would still contribute as an instance to the total count for that day. It is essential to consider that students enrolled for various years are counted in the annual participation count. Their daily interactions are recorded for each year they remain active, which could contribute to the

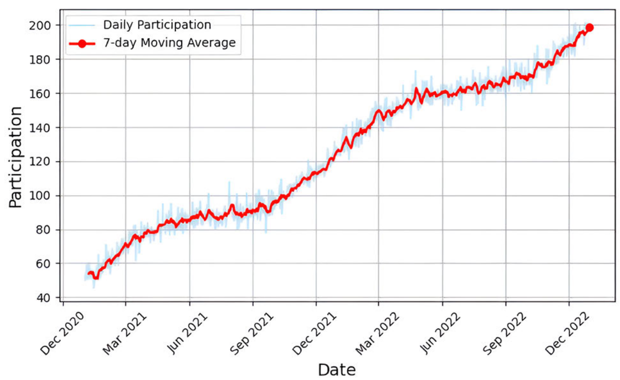

FIGURE 7. Two-year trend of student engagement in online learning.

TABLE 10. Model performance and recommendation system.

Model	precision	Recall	F1-score
Case 1: ARIMA	0.85	0.78	0.81
Case 2: Prophet	0.92	0.88	0.90
Case 3: K-Means	0.72	0.68	0.70
Case 4: DBSCAN	0.65	0.75	0.70
Recommendation System	0.88	0.90	0.89

apparent increase in engagement from year to year. We have carefully distinguished between continued long-term student engagement and genuine overall platform user engagement increases.

The increase from 45 daily interactions in December 2020 to 200 in December 2022 corresponds to student retention across multiple terms and the overall increase in the user base due to new enrollments. The pattern of seasonality, indicated by peaks and valleys, corresponds to the rhythm of the academic calendar, with notable decreases during vacation periods and increases around testing periods or other significant academic events. Participation trends are essential to inform the educational institution's resource planning and intervention strategies. They signal periods of high demand on the online system and help identify optimal times to schedule maintenance or introduce new functionality.

3) MODEL EVALUATION

To evaluate the performance of the models used in the student participation monitoring system, key metrics such as precision, completeness, and F1-score were employed to analyze their effectiveness in detecting patterns of student participation. The results are presented in Table 10.

From the data obtained, it has been identified that the ARIMA model demonstrated solid performance in terms of precision, with a value of 0.85. This means that the predictions made by ARIMA coincided with the observed data in 85% of the cases. Furthermore, the model showed good recall capacity (0.78) and an F1-score of 0.81, indicating that it effectively identified short-term participation patterns. ARIMA excelled at modeling short-term participation trends, which is crucial for immediate decision-making.

TABLE 11. Comparison of student participation before and after implementation of the monitoring system.

Before Control	After Control	Before Experimental	After Experimental
54.97	36.85	53.58	45.71
48.62	46.79	55.61	48.40
56.48	47.57	60.83	61.47
65.23	42.98	60.54	60.10
47.66	49.39	36.22	53.79

Note: "Control" refers to the group that did not use the monitoring system, while "Experimental" refers to the group that used the system. Each cell represents the average participation rate.

On the other hand, the Prophet model achieved exceptional performance in all metrics. With a precision of 0.92, a recall of 0.88, and an F1-score of 0.90, this model proved highly effective in predicting long-term trends and detecting seasonal patterns. Its predictions fit well with variations in student engagement data, supporting its ability to provide reliable and accurate information for long-term decision-making.

In K-Means' case, reasonable performance is observed with a precision of 0.72, a recall of 0.68, and an F1-score of 0.70. K-Means was used to segment students into groups with similar participation profiles. Although the results are acceptable, the precision of the segmentation could be improved.

DBSCAN proved effective in identifying groups of students with unconventional participation profiles. However, its performance is reflected in a precision of 0.65, a recall of 0.75, and an F1-score of 0.70. These values suggest that while DBSCAN can identify exciting clusters, there may be some overlap or noise in the results.

In addition to the time series and clustering models, a recommendation system was implemented to personalize the student experience. The recommendation system showed a precision of 0.88, a completeness (recall) of 0.90, and an F1-score of 0.89. This indicates that the recommender system effectively provides relevant suggestions and improves student engagement.

Model evaluation reveals that Prophet is best at long-term prediction and seasonal pattern detection, while ARIMA is effective at short-term trend modeling. Clustering models like K-Means and DBSCAN have room for improved segmentation precision. Furthermore, the personalized recommendation system shows strong performance in improving student engagement.

4) RESULTS AND SYSTEM IMPACT ANALYSIS

To effectively evaluate the impact of the student participation monitoring system, statistical analyses were performed using data collected from 100 students Table 11, divided into two groups: an experimental group, which used the monitoring system, and a control group, which he didn't. Two key metrics were analyzed before and after the implementation of the system: Participation (P) and Academic Performance (AR).

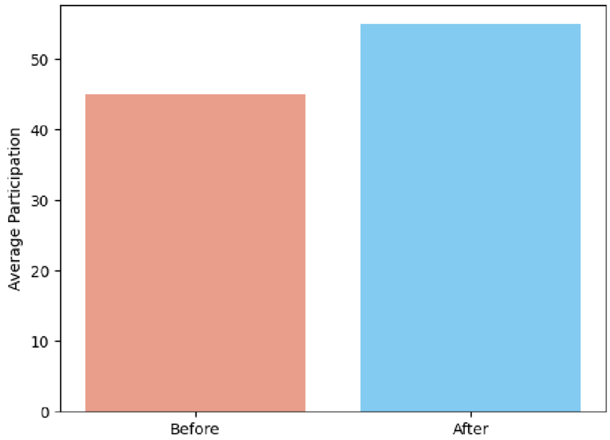

FIGURE 8. Comparison of student participation before and after implementation of the monitoring system.

Participation was measured as the percentage of activities completed in the course, while Academic Performance was evaluated from the average scores on the assessments. A t-test was carried out to compare the means of participation in the experimental group before and after the implementation of the system.

The t-test results showed a statistically significant difference in student participation, with a statistical value of −3.16 and a p-value of approximately 0.0021. This result indicates a substantial increase in student participation in the experimental group after implementing the monitoring system.

Figure 8 compares mean participation before and after implementation in the experimental group. An increase in average participation is observed, which supports the effectiveness of the implemented system in improving student participation. These findings suggest that the monitoring system positively and significantly impacts student participation in course-related activities.

The results indicate that the student participation monitoring system is an effective tool to improve student participation and academic performance, supporting its implementation in educational settings to enhance student participation and success.

IV. DISCUSSION

The transition from traditional in-person education to virtual environments presents unique challenges impacting student engagement. Factors such as distractions at home, work and family responsibilities, and the need for self-management and motivation emerge as critical elements influencing students' engagement with course material and learning activities. Our study recognizes the importance of these external factors and focuses on measuring their impact on student engagement by analyzing data collected in the tracking system. By correlating engagement patterns with these factors, we seek to understand how pedagogical strategies and practices in online education can be improved to address these challenges and enhance student learning effectively.

This data analysis has uncovered substantial variability in student engagement over time, which resonates with earlier studies [45] that documented similar patterns. The observed decline in engagement during vacation periods is worth noting, aligning with existing literature suggesting that students tend to disengage during academic breaks [46].

The pivotal aspect of our results lies in evaluating forecasting models, including ARIMA and Prophet, concerning their effectiveness in predicting student engagement. These findings echo previous research assessing the applicability of such techniques within educational contexts [47], [48]. By doing so, our study sheds light on the utility of these models as invaluable tools for anticipating student engagement, which, in turn, can inform strategic decision-making in educational institutions to enhance student retention and academic performance.

This work contributes to the broader knowledge of tracking online student engagement and reinforces many established findings. For instance, the documented variability in student engagement over time corroborates the validity of our data and analyses, consistent with prior research [49]. Similarly, the results regarding the effectiveness of forecasting models align with previous studies that have evaluated the suitability of these techniques in educational settings. However, our study goes beyond mere replication; it makes several notable contributions to online education and student engagement tracking [50], [51].

One of the most noteworthy contributions is our study's potential to enhance student retention and academic performance in online learning environments. Institutions can proactively identify and assist at-risk students by recognizing participation patterns and impacting their learning journey [33]. This research serves as a solid foundation for future investigations in online education. Identified areas, such as applying forecast models and understanding seasonal participation patterns, could be subjects of more in-depth and specific research endeavors [52], [53].

In evaluating the machine learning models used to monitor student engagement, key metrics such as accuracy, completeness, and F1 score were applied to analyze their effectiveness in detecting patterns of student engagement. It is essential to specify that the precision reported here is of the macro type, which considers equality of conditions for each class of student participation, regardless of its frequency in the data set.

The choice to use macro rather than micro precision is based on our study's goal of treating all types of student engagement with equal importance. This is crucial to the educational goal of providing equitable support to all students. This decision aligns with the study's objectives of ensuring an inclusive and representative assessment of student participation in the online learning platform.

Additionally, our accuracy metrics were compared with similar studies in the field of online education, which generally report accuracies in a similar range. This validates the appropriateness of our precision levels in relation to industry regulations and reinforces the relevance of our findings in the broader context of educational research.

It is essential to recognize the limitations of our study. First, the data are derived from a single online educational institution, which may affect the generalizability of the results. Future research should attempt to replicate this study in multiple online educational environments to address this limitation and validate the findings more comprehensively [54]. Additionally, while we evaluated the effectiveness of various forecasting models, we may not have considered all available techniques.

Through the participation tracking system, our study delves into the comments and opinions shared by students and educators. These qualitative insights enrich our understanding of participants' experiences and illuminate the system's effectiveness in promoting engagement and improving academic performance. Qualitative analysis of this feedback revealed several key themes and observations. Study participants provided valuable feedback on the usefulness of the system and its influence on their participation in online courses. Observations highlighted the system's ability to foster more profound interaction with course content and facilitate progress tracking. Some students expressed that personalized feedback helped them identify areas for improvement and adjust their study habits. In contrast, others noted that the system reminded them of the importance of active participation.

Educators who used the system shared their observations of its ability to identify students who require additional support. They also noted that real-time student engagement tracking allowed them to adapt their teaching approach and provide personalized guidance when necessary. Several educators praised the system's seamless integration into their workflow and its effectiveness in improving student communication.

V. CONCLUSION

This work has provided a clear view of student engagement in the online educational environment, supported by data analysis and evaluation of forecasting models. The results identified seasonal patterns and trends in student engagement, shedding light on the unique dynamics of online education.

The developed method allows for the documentation of significant variability in student participation throughout the different stages of the academic year. This includes identifying seasonal patterns, such as lower participation during holiday periods. These findings support the validity and consistency of our data and corroborate previous research suggesting that students may disengage during academic breaks. This direct correlation between our data and previously documented trends underscores the findings' relevance and applicability in real educational contexts.

This work has evaluated the effectiveness of forecasting models, such as ARIMA and Prophet, in accurately predicting student engagement. The results confirm the usefulness of these tools in educational contexts, allowing institutions to anticipate participation trends and take proactive measures. The accuracy of these models, detailed in Section IV, shows how statistical analyses support these models' application in academic and administrative planning.

Educational institutions can use the insights generated by this study to make informed decisions and improve student retention and academic performance in online environments. By implementing forecasting models and understanding engagement patterns, institutions can more effectively support their students, which, in turn, fosters academic success. The results show how variations in the implementation of teaching strategies can directly influence participation and retention rates.

The results obtained lay a solid foundation for future research on online education and monitoring student participation. Areas of interest have been identified, such as exploring new forecasting methodologies and approaches and a deeper understanding of student engagement patterns. These areas may be the focus of more detailed and specific future research.

Our findings suggest that future research could explore additional forecasting models to monitor online student engagement. For example, the applicability of more advanced machine learning models could be investigated to improve prediction accuracy. This reflects an opportunity to integrate innovative approaches that directly address the changing dynamics of online education.

Another research topic that can be addressed is the analysis of retention factors. For this reason, the importance of investigating factors influencing student retention in online educational environments has been highlighted. This could include analysis of additional variables, such as the quality of course content, student-instructor interaction, and other factors that may influence student engagement.

REFERENCES

[1] S. Shehzadi, D. S. Jamshed, H. Nosheen, J. Asim, S. A. Bhatti, and S. Akram, "The impact of E-service quality, Facebook usage and artificial intelligence on E-learning systems: The mediating role of student satisfaction," *Int. J. Manage. Res. Emerg. Sci.*, vol. 11, no. 1, pp. 1–12, Mar. 2021, doi: 10.56536/ijmres.v11i1.126.

[2] A. Ferrario, M. Loi, and E. Viganò, "In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions," *Philosophy Technol.*, vol. 33, no. 3, pp. 523–539, Sep. 2020, doi: 10.1007/s13347-019-00378-3.

[3] L. A. Hildebrand, C. J. Pierce, M. Dennis, M. Paracha, and A. Maoz, "Artificial intelligence for histology-based detection of microsatellite instability and prediction of response to immunotherapy in colorectal cancer," *Cancers*, vol. 13, no. 3, p. 391, Jan. 2021, doi: 10.3390/cancers13030391.

[4] L. F. Daghestani, L. F. Ibrahim, R. S. Al-Towirgi, and H. A. Salman, "Adapting gamified learning systems using educational data mining techniques," *Comput. Appl. Eng. Educ.*, vol. 28, no. 3, pp. 568–589, May 2020, doi: 10.1002/cae.22227.

[5] A. Adib, A. Zaerpour, and M. Lotfirad, "On the reliability of a novel MODWT-based hybrid ARIMA-artificial intelligence approach to forecast daily snow depth (Case study: The western part of the rocky mountains in the U.S.A)," *Cold Regions Sci. Technol.*, vol. 189, Sep. 2021, Art. no. 103342, doi: 10.1016/j.coldregions.2021.103342.

[6] Y. Duan, H. Wang, M. Wei, L. Tan, and T. Yue, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction," *Petroleum*, vol. 8, no. 2, pp. 270–277, Jun. 2022, doi: 10.1016/j.petlm.2021.09.001.

[7] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data Brief*, vol. 29, Apr. 2020, Art. no. 105340, doi: 10.1016/j.dib.2020.105340.

[8] C. C. Chen, Y. Chen, and C. Y. Kang, "Estimating student-teacher ratio with ARIMA for primary education in fluctuating enrollment," *ICIC Exp. Lett., Part B, Appl.*, vol. 12, no. 6, pp. 1–19, 2021, doi: 10.24507/icicelb.12.06.515.

[9] M. Elhossiny, R. Eladly, and A. Saber, "The integration of psychology and artificial intelligence in e-learning systems to guide the learning path according to the learner's style and thinking," *Int. J. Adv. Appl. Sci.*, vol. 9, no. 12, pp. 162–169, Dec. 2022, doi: 10.21833/ijaas.2022.12.020.

[10] S. H. Hessen, H. M. Abdul-kader, A. E. Khedr, and R. K. Salem, "Developing multiagent E-learning system-based machine learning and feature selection techniques," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–8, Jan. 2022, doi: 10.1155/2022/2941840.

[11] F. Gazzawe, M. Mayouf, R. Lock, and R. Alturki, "The role of machine learning in E-learning using the web and AI-enabled mobile applications," *Mobile Inf. Syst.*, vol. 2022, pp. 1–10, Jun. 2022, doi: 10.1155/2022/3696140.

[12] D. Azzolini, S. Marzadro, E. Rettore, K. Engelhardt, B. Hertz, and P. Wastiau, "Raising teacher retention in online courses through personalized support. Evidence from a cross-national randomized controlled trial," *J. Res. Educ. Effectiveness*, vol. 16, no. 2, pp. 300–325, Apr. 2023, doi: 10.1080/19345747.2022.2100850.

[13] P. Bawa, *The Importance of Studying Retention Issues in Online Courses Retention in Online Courses: Exploring Issues and Solutions-A Literature Review*. Newbury Park, CA, USA: Sage, 2016.

[14] T. Susnjak, "Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT," *Int. J. Artif. Intell. Educ.*, vol. 1, pp. 1–31, Jun. 2023, doi: 10.1007/s40593-023-00336-3.

[15] G. Morales-Romero, N. Trinidad-Loli, B. Caycho-Salas, Y. Paucar-Manrique, C. León-Velarde, S. Gamarra-Mendoza, N. Alvarado-Bravo, A. Torrez-Quiroz, C. Aliaga-Valdez, and F. Aldana-Trejo, "Perception of teaching performance in the virtual learning environment," *Int. J. Eval. Res. Educ.*, vol. 10, no. 4, p. 1221, Dec. 2021, doi: 10.11591/ijere.v10i4.22056.

[16] Z. Zhang, "A method of recommending physical education network course resources based on collaborative filtering technology," *Sci. Program.*, vol. 2021, pp. 1–9, Oct. 2021, doi: 10.1155/2021/9531111.

[17] R. Aguayo, C. Lizarraga, and Y. Quiñonez, "Evaluación del desempeño académico en entornos virtuales utilizando el modelo PNL," *Revista Ibérica de Sistemas e Tecnologias de Informação*, vol. 2021, no. 41, pp. 34–49, Mar. 2021.

[18] L. A. Costa, L. M. P. Sanches, R. J. R. Amorim, L. D. N. Salvador, and M. V. D. S. Souza, "Monitoring academic performance based on learning analytics and ontology: A systematic review," *Informat. Educ.*, vol. 19, pp. 361–397, Sep. 2020, doi: 10.15388/infedu.2020.17.

[19] S. Garg, T. Bag, and A. Mitschele-Thiel, "Decentralized machine learning based network data analytics for cognitive management of mobile communication networks," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp.*, Apr. 2022, pp. 1–9, doi: 10.1109/NOMS54207.2022.9789936.

[20] V. E. D. Mondini and M. J. C. D. S. Domingues, "Gestão da retenção de alunos em cursos on-line sob a perspectiva da aceitação da tecnologia," *Revista Brasileira de Educação*, vol. 23, Sep. 2018, Art. no. e230050, doi: 10.1590/s1413-24782018230050.

[21] J. E. Scarpin, V. E. Dagnoni Mondini, and M. R. S. Scarpin, "Technology acceptance factors and student retention in online courses," *E-J. Bus. Educ. Scholarship Teaching*, vol. 12, no. 3, pp. 44–68, 2018.

[22] I. El Naqa and M. J. Murphy, "What Is Machine Learning?", in *Machine Learning in Radiation Oncology*. I. El Naqa, R. Li, M. Murphy, Eds. Cham. Switzerland: Springer, 2015, doi: 10.1007/978-3-319-18305-3_1.

[23] M. B. Yel, M. Zarlis, S. Suwilo, and S. Efendi, "Framework model smart E-learning berbasis kecerdasan buatan," *Prosiding Seminar Nasional Riset Inf. Sci.*, vol. 4, no. 2, pp. 119–127, Jul. 2022.

[24] J. D. Cochran, S. M. Campbell, H. M. Baker, and E. M. Leeds, "The role of student characteristics in predicting retention in online courses," *Res. Higher Educ.*, vol. 55, no. 1, pp. 27–48, Feb. 2014, doi: 10.1007/s11162-013-9305-8.

[25] N. Sellami, B. Toven-Lindsey, M. Levis-Fitzgerald, P. H. Barber, and T. Hasson, "A unique and scalable model for increasing research engagement, STEM persistence, and entry into doctoral programs," *CBE—Life Sci. Educ.*, vol. 20, no. 1, pp. 1–12, Mar. 2021, doi: 10.1187/cbe.20-09-0224.

[26] M. C. Sáiz-Manzanares, R. Marticorena-Sánchez, L. J. M. Antón, I. González-Díez, and M. Á. C. Martín, "Using eye tracking technology to analyse cognitive load in multichannel activities in university students," *Int. J. Hum.–Comput. Interact.*, vol. 40, pp. 1–19, Mar. 2023, doi: 10.1080/10447318.2023.2188532.

[27] R. J. Ruhf, C. T. Williams, M. Zelinsky, and L. W. Becho, "Barriers to collecting student participation and completion data for a national STEM education grant program in the United States: A multiple case study," *Int. J. STEM Educ.*, vol. 9, no. 1, pp. 1–18, Dec. 2022, doi: 10.1186/s40594-022-00348-w.

[28] M. Husni, "Integration of the civil society development of the prophet muhammad," *Pappaseng, Int. J. Islamic Literacy Soc.*, vol. 1, no. 1, pp. 47–57, Apr. 2022, doi: 10.56440/pijils.v1i1.3.

[29] U. Khan, "An ARIMA forecasting approach to evaluate the scenario of male and female higher education," *Int. J. Entrepreneurship*, vol. 25, no. 4, pp. 1–7, 2021.

[30] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *Int. J. Eng. Bus. Manage.*, vol. 10, Jan. 2018, Art. no. 184797901880867, doi: 10.1177/1847979018808673.

[31] Z. Wu, M. Guan, M. Liu, L. Xia, C. Lv, X. Cao, and H. Chen, "Research on intelligent scheduling algorithm of high altitude platform system for wide area Internet of Things," *Wireless Netw.*, vol. 30, pp. 1–7, Nov. 2021, doi: 10.1007/s11276-021-02842-5.

[32] Y. Guan, Q. Wei, and G. Chen, "Deep learning based personalized recommendation with multi-view information integration," *Decis. Support Syst.*, vol. 118, pp. 58–69, Mar. 2019, doi: 10.1016/j.dss.2019.01.003.

[33] J. Xiao, M. Wang, B. Jiang, and J. Li, "A personalized recommendation system with combinational algorithm for online learning," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 3, pp. 667–677, Jun. 2018, doi: 10.1007/s12652-017-0466-8.

[34] W. Villegas-Ch, A. Mera-Navarrete, and J. García-Ortiz, "Data analysis model for the evaluation of the factors that influence the teaching of university students," *Computers*, vol. 12, no. 2, p. 30, Jan. 2023, doi: 10.3390/computers12020030.

[35] R. Hao, L. Zhang, J. Liu, Y. Liu, J. Yi, and X. Liu, "A promising approach: Artificial intelligence applied to small intestinal bacterial overgrowth (SIBO) diagnosis using cluster analysis," *Diagnostics*, vol. 11, no. 8, p. 1445, Aug. 2021, doi: 10.3390/diagnostics11081445.

[36] A. Mele, A. Vitiello, M. Bonano, A. Miano, R. Lanari, G. Acampora, and A. Prota, "On the joint exploitation of satellite DInSAR measurements and DBSCAN-based techniques for preliminary identification and ranking of critical constructions in a built environment," *Remote Sens.*, vol. 14, no. 8, p. 1872, Apr. 2022, doi: 10.3390/rs14081872.

[37] W. Villegas-Ch., J. García-Ortiz, I. Urbina-Camacho, and A. Mera-Navarrete, "Proposal for a system for the identification of the concentration of students who attend online educational models," *Computers*, vol. 12, no. 4, p. 74, Apr. 2023, doi: 10.3390/computers12040074.

[38] D. Kc, "Evaluation of moodle features at Kajaani university of applied sciences—Case study," *Proc. Comput. Sci.*, vol. 116, pp. 121–128, Jun. 2017, doi: 10.1016/j.procs.2017.10.021.

[39] H.-T. Lee, J.-S. Lee, H. Yang, and I.-S. Cho, "An AIS data-driven approach to analyze the pattern of ship trajectories in ports using the DBSCAN algorithm," *Appl. Sci.*, vol. 11, no. 2, p. 799, Jan. 2021, doi: 10.3390/app11020799.

[40] M. Li, H. Wu, and H. Zhang, "Matrix factorization for personalized recommendation with implicit feedback and temporal information in social ecommerce networks," *IEEE Access*, vol. 7, pp. 141268–141276, 2019, doi: 10.1109/ACCESS.2019.2943959.

[41] D.-F. Chang and H.-C. ChangTzeng, "Patterns of gender parity in the humanities and STEM programs: The trajectory under the expanded higher education system," *Stud. Higher Educ.*, vol. 45, no. 6, pp. 1108–1120, Jun. 2020, doi: 10.1080/03075079.2018.1550479.

[42] D. L. Wijaya and A. Maryani, "Efektivitas komunikasi dalam pembelajaran berbasis E-learning di universitas Islam Bandung," in *Proc. Bandung Conf. Series*, 2023, vol. 3, no. 1, pp. 289–297, doi: 10.29313/bcscm.v3i1.5854.

[43] S.-C. Necula and V.-D. Păvăloaia, "AI-driven recommendations: A systematic review of the state of the art in e-commerce," *Appl. Sci.*, vol. 13, no. 9, p. 5531, Apr. 2023, doi: 10.3390/app13095531.

[44] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–19, May 2021, doi: 10.1007/s42979-021-00592-x.

[45] Y. Lin and Z. Yu, "A bibliometric analysis of artificial intelligence chatbots in educational contexts," *Interact. Technol. Smart Educ.*, vol. 21, no. 2, pp. 189–213, Apr. 2024, doi: 10.1108/itse-12-2022-0165.

[46] P. D. Bons, T. de Riese, S. Franke, M.-G. Llorens, T. Sachau, N. Stoll, I. Weikusat, J. Westhoff, and Y. Zhang, "Comment on 'Exceptionally high heat flux needed to sustain the Northeast Greenland Ice Strea,' by Smith-Johnsen et al. (2020)," *Cryosphere*, vol. 15, no. 5, pp. 2251–2254, May 2021, doi: 10.5194/tc-15-2251-2021.

[47] D. K. DiGiacomo, "Digital civic learning in schools: Youth perspectives and experiences," *Inf. Learn. Sci.*, vol. 122, nos. 11–12, pp. 709–725, Nov. 2021, doi: 10.1108/ils-01-2020-0013.

[48] D. F. Chang, E. Y. Liu, and B. Y. Chen, "Heterogeneity of students program participation in expanding higher education interpreted by the blau index and ARIMA," *ICIC Exp. Lett., Part B, Appl.*, vol. 14, no. 4, pp. 1–20, 2023, doi: 10.24507/icicelb.14.04.349.

[49] L. Zhu, J. Shen, G. Gartner, and Y. Hou, "Personalized landmark sequence recommendation method using LSTM-based network for navigating in large hospitals," *Abstr. Integr. Comput.-Aided Eng.*, vol. 3, pp. 1–2, Dec. 2021, doi: 10.5194/ica-abs-3-328-2021.

[50] M. Abubakar, Y. Che, L. Ivascu, F. M. Almasoudi, and I. Jamil, "Performance analysis of energy production of large-scale solar plants based on artificial intelligence (machine learning) technique," *Processes*, vol. 10, no. 9, p. 1843, Sep. 2022, doi: 10.3390/pr10091843.

[51] S. A. Shahriar, I. Kayes, K. Hasan, M. Hasan, R. Islam, N. R. Awang, Z. Hamzah, A. E. Rak, and M. A. Salam, "Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric $PM_{2.5}$ forecasting in Bangladesh," *Atmosphere*, vol. 12, no. 1, p. 100, Jan. 2021, doi: 10.3390/atmos12010100.

[52] A. Gorlapalli, S. Kallakuri, P. D. Sreekanth, R. Patil, N. Bandumula, G. Ondrasek, M. Admala, C. Gireesh, M. S. Anantha, B. Parmar, B. K. Yadav, R. M. Sundaram, and S. Rathod, "Characterization and prediction of water stress using time series and artificial intelligence models," *Sustainability*, vol. 14, no. 11, p. 6690, May 2022, doi: 10.3390/su14116690.

[53] G. Chitikela, M. Admala, V. K. Ramalingareddy, N. Bandumula, G. Ondrasek, R. M. Sundaram, and S. Rathod, "Artificial-intelligence-based time-series intervention models to assess the impact of the COVID-19 pandemic on tomato supply and prices in Hyderabad, India," *Agronomy*, vol. 11, no. 9, p. 1878, Sep. 2021, doi: 10.3390/agronomy11091878.

[54] V. R. More, G. Singh, and K. Patwardhan, "Introducing hybrid problem-based learning modules in ayurveda education: Results of an exploratory study," *J. Alternative Complementary Med.*, vol. 26, no. 2, pp. 130–137, Feb. 2020, doi: 10.1089/acm.2019.0293.

WILLIAM VILLEGAS-CH (Member, IEEE) received the master's degree in communications networks and the Ph.D. degree in computer science from the University of Alicante. He is a Professor of information technology with the Universidad de Las Américas, Quito, Ecuador. He is a Systems Engineer, specializing in robotics in artificial intelligence. He has participated in various conferences as a speaker on topics, such as ICT in education and how they improve educational quality and student learning. His main articles focus on the design of ICT systems, models, and prototypes applied to different academic environments, especially with the use of big data and artificial intelligence as a basis for creating intelligent educational environments. His main research interests include web applications, data mining, and e-learning.

JOSELIN GARCÍA-ORTIZ is currently pursuing the degree in cybersecurity engineering with the Universidad de Las Américas. She is a Software Engineer with Capmation Inc., Quito, Ecuador, where she applies her knowledge in software development and cybersecurity. She has a strong interest in the intersection of AI and its applications to education and has previous experience in researching on the subject.

SANTIAGO SÁNCHEZ-VITERI was born in Quito, Ecuador, in 1984. He received the System Engineering degree from Salesiana Polytechnic University (UPS), Ecuador, in 2017, where he is currently pursuing the master's degree in telematics management. Since 2010, he has been working in the area of telecommunications and networking of IT with Universidad Internacional del Ecuador. He has been a Professor of computer science with Universidad Internacional del Ecuador, where he is also an Administrator of computer servers and internet connection equipment and the Moodle and CANVAS learning management systems. He has participated in more than ten indexed articles on telecommunications, big data, and innovation.