International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2
DOI: https://doi.org/10.17762/ijritcc.v11i2.6108
***Article Received**: 10 December 2022 **Revised**: 15 January 2023 **Accepted**: 27 January 2023*

Student Engagement Prediction in Online Session

Y. Trilochan Sashank[1], Dr. V. Kakulapati[2], Dr. Sunil Bhutada[3]
[1]Sreenidhi Institute of Science and Technology,
Yamnampet, Ghatkesar, Hyderabad, Telangana-501301.
sharmasashi287@gmail.com,
[2]Sreenidhi Institute of Science and Technology,
Yamnampet, Ghatkesar, Hyderabad, Telangana-501301.
vldms@yahoo.com,
[3]Sreenidhi Institute of Science and Technology,
Yamnampet, Ghatkesar, Hyderabad, Telangana-501301.
sunilb@sreenidhi.edu.in

Abstract: The individuals who make up the globe constantly advance into the future and improve both their personal lives and the conditions in which they live. One ‘s education is the basis of one ‘s knowledge. Humans' education has a significant impact on their behavior and IQ. Through the use of diverse pedagogical techniques, instructors always play a part in changing students' ways of thinking and developing their social and cognitive abilities. However, getting students to participate in an online class is still difficult. In this study, we created an intelligent predictive system that aids instructors in anticipating students' levels of interest based on the information they learn in an online session and in motivating them through regular feedback. The level of students' engagement is divided into three tiers based on their online session activities (Not engaged, passively engaged, and actively engaged). The given data was subjected to the application of Decision Trees (DT), Random Forest Classifiers (RF), Logistic Regression (LR), and Long Short-Term Memory Networks are among the numerous machine learning approaches (LSTM). According to performance measurements, LSTM is the most accurate machine learning algorithm. The instructors can get in touch with the students and inspire them by improving their teaching approaches based on the results the system produces.

Keywords: student, IQ, education, instructors, machine learning, algorithm, teaching, session, prediction, LSTM.

I. INTRODUCTION:

The rapid digitization of educational institutions is having a major impact on how teachers and other educational staff can facilitate data collection. In addition to offline-to-online learning styles and changes in institutional governance, the wealth of data about educational institutions is also changing very rapidly. A few examples of the numerous ways that web-based learning is currently used extensively in education (LMS) include Massive Open Online Courses (MOOCs), Virtual Learning Environments (VLEs), and Learning Management Systems. MOOC students can study anytime, anywhere [1]. MOOCs offer a new way of teaching, changing the traditional way of learning and attracting learners from all over the world. The three most popular platforms are Harvard, Edx, and Coursera. MOOCs also contribute to the development of higher education [2].

However, it is challenging for teachers to keep track of everyone and identify disinterested students due to a large number of students in the class. You need an intelligent system that can analyze academic data in e-learning system reports or logs to predict student participation. Because student engagement is a measure of academic performance, it is essential to identify students who lack motivation and work to increase their engagement [3, 4]. A lot of people think that student engagement [5] is a proximal outcome that leads to distal outcomes like higher academic performance and lower rates of dropout. Provide a suggestion for a predictive method that is effective for both students. The authors suggest predicting the level of student engagement by using student data activity from the LMS (attendance at meetings, participation in forums and groups, access to course materials, etc.). All of this data can be found in a lot of reports. The data from the author's Kaggle are used in this study. It provides a variety of reports, including:

(a) Outline of course activities;
(b) Summary of session activities
(c) Student profiles for individual courses;
(e) All user activity in content sections;
(d) A general breakdown of user activity;
(f) Participation in forums;
(g) Participation in groups;
(h) Completion of courses;
(i) Completion of course coverage reports; and
(j) Participation in courses by users.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2
DOI: https://doi.org/10.17762/ijritcc.v11i2.6108
***Article Received**: 10 December 2022 **Revised**: 15 January 2023 **Accepted**: 27 January 2023*

1.1. Objectives*:*

Can student involvement be classified as Active (AE), Passive (PE), or Non-Involved (NE) based on the amount of information learned via online sessions?

Among LR, DT, LSTM, and random forest classifiers, which is the best machine learning classification algorithm in the proposed study? Try to predict the level of involvement. Depending on the student's activity in the online course, various Students are tested using machine learning algorithms [6], and they are placed into one of her three categories:

N, E, or A. A novel labeling system is used to choose these categories, and it takes teacher approval and student performance into account (i.e. GPA or percentile). In order to forecast the degree of student engagement, the suggested method makes use of the most precise prediction models in terms of accuracy, precision, recall, and F1 score requirements. Best practices implementation can boost student engagement and enhance academic performance and achievement. According to the synthetic data from which the LMS reports were gathered, this study suggested dividing students into three main categories (actively engaged, passively involved, and not engaged), respectively. It should be noted that this study differs from many others in that it emphasizes system capabilities. Various.

1.2. Motivation for the study

Everywhere in the world, on-campus education is practiced the conventional way. Since the previous decade, web-based learning has grown significantly in popularity, and many institutions have begun to employ it, particularly during the lockdown brought on by the COVID-19 epidemic. Through their pedagogical methods, educators can instruct their pupils online. However, it is still difficult for teachers to assess and forecast students' levels of participation. Student engagement is a complex and diverse phenomenon with behavioral, components that are cognitive, social, and emotional [11], [12]. The final element of student engagement is crucial yet difficult to achieve, particularly in an online setting. where the emotional bond between students and professors is difficult to see and assess. Student engagement may be compromised. As a result, teachers need to improve their work and further develop teaching methods (animation of presentations, pop-up quizzes during lectures, etc.). This survey helps predict student engagement based on student comprehension and time spent in online sessions. This benefits both instructors and students. Such a system allows teachers to change pedagogical approaches and course materials when it is clear that students are not working hard enough. Giving students regular comments and encouragement can also increase student engagement. Despite the structural relevance Few academics have attempted to establish useful and trustworthy indicators of student involvement that are pertinent to this setting.

II. RELATED WORK

Numerous studies show a positive correlation between student participation and course outcomes. For instance, Atherton [7] demonstrated that students who regularly take assessments and obtain study materials through web-based systems perform well on exams. According to an additional study, students who are very engaged likely to perform well on course quizzes and evaluations. [8] According to Rodgers [9], there was a strong correlation between student use of an e-learning system and the results of the courses. However, the majority of earlier research on engagement has overlooked student participation in web-based systems in favor of conventional instruction in colleges and schools. Furthermore, earlier studies on student engagement have relied on a survey, qualitative, and statistical methodologies; however, these statistical tools are unable to detect hidden knowledge in student data.

Additionally, statistically qualitative techniques lack scalability and generalizability. Surveys are a poor choice for gauging student interest because, for instance, smaller kids cannot grasp the questions, and they take a long time to complete them. The fact that these studies are focused on the course and the student's emotions and behavioral structure is another drawback. However, student engagement can also be influenced by students' involvement in educational activities.

A current study uses VLE log data to predict students' low participation in web-based learning systems. Students are not interrupted when accessing log data using machine learning techniques [10] and the data is not time-sensitive.

III.METHODOLOGY

The most popular methods for gauging student participation are surveys and questionnaires, as well as using information that the students themselves submit. Case studies, observations, checklists, and evaluation scales filled out by teachers are some other strategies [13]. It is crucial to specify this breadth of involvement in order to accurately assess student participation in a particular setting [14], [15]. Based on student activity and behavior in courses from LMS data, this article illustrates the dynamic nature of the behavioral, social, and cognitive components of student involvement.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2
DOI: https://doi.org/10.17762/ijritcc.v11i2.6108
***Article Received**: 10 December 2022 **Revised**: 15 January 2023 **Accepted**: 27 January 2023*

Based on the amount of information learned in online sessions, is it possible to classify a student's degree of participation as Actively Engaged (AE), Passively Engaged (PE), or Not Engaged (NE)? I'll start by responding to that query. To divide student interaction into three categories, we created a methodology. This strategy is based on the student's percentile on the exam and the professor's praise [16, 17]. We used a categorical goal variable (class) to display the pupils' proportion of knowledge. Different percentile ranges are represented by the three categories AE, PE, and NE. Based on suggestions and acceptance from the instructor, the resulting student engagement levels were then updated and validated. The labeling stage in the record preparation procedure is the outcome of this step. In order to respond to the second query

Among DT, LR, LSTM, and random forest classifiers, which machine learning classification algorithm performs best in your proposed research?) For applicability to target applications, four machine learning We chose models (DT, LR, LSTM, and Random Forest). study. A supervised learning classifier that can be used to classify and predict categorical variables is the Decision Tree Algorithm or DT for short**.**

Both continuous variables for regression and categorical variables for classification are supported by the Random Forest Classifier. It gives excellent results when it comes to classification problems. Recurrent Neural Network (RNN) technology, known as LSTM, is well-suited for large data sets, especially those with a large number of features. The outcome of the current input depends on the results of earlier calculations. As a result, the RNN is iterative because it completes the same task for every data input. It produces, copies, and sends outputs to the recurrent network. When making judgments, it takes into account both the current input and the outcome learned from earlier inputs. Logistic regression is used to forecast the output for categorical dependent variables. As a result, the outcomes must be categorical or discrete. It gives probability values in the range of 0 to 1, rather than exact values between 0 and 1. can be true or false, 0 or 1, yes or no, and so forth. We calculated the recall, accuracy, precision, and F1 score for these four selected models, and we compared their performance metrics.

As previously stated, our objective is to forecast whether the student would participate at an AE, PE, or NE level. The proposed predictive system aims to measure student interest and inspire them to consistently advance and switch between categories. This is because student performance is directly impacted by involvement. Additionally, it enables motivated pupils to keep up their current rate of work.

IV. IMPLEMENTATION ANALYSIS

Data description. Data for this study was gathered from a variety of internet sources (including Kaggle, GitHub, etc.), and many rows and columns were left out to ensure the study's results were as accurate as possible. So, we ended up with a dataset that included 10 columns and 100 data rows. The first column includes the student's ID, and the remaining columns include information about engagement factors like the amount of time spent by the user preparing the object materials, reviewing the study materials, spending time on related materials to accomplish a task, completing the task using related materials, completing the task successfully, and the level of engagement.

Dataset labeling. Before creating predictive models, we classified a student's participation into the AE, PE, and NE groups. It's challenging to evaluate a student's understanding during an online session. Therefore, we must create our labeling technique. Performance in the relevant subjects can be used to determine the student's level of understanding. The instructor's assessment of the student's performance is based on the percentile they achieved. Therefore, the level of student involvement is determined based on how much knowledge students learned in an online session and how long they spent in an online session.

Models building and evaluation. After collecting and pre-processing the data, We developed four forecasting models (LR, DT, LSTM, and Random Forest). Eleven commitment traits are employed as input traits, and the goal variable is the anticipated ultimate student performance (percentile). In the 10-way cross-validation procedure used to create each model, nine subsets are utilized for training, and the last subset is used for testing. The performance accuracy of machine learning models is assessed using the metrics listed below. Classifier Predictive Power is an abbreviation. accuracy serving as a gauge of the classifier's performance. Classifier Sensitivity's acronym. A balance between recall and accuracy can be seen in the F1 score.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2
DOI: https://doi.org/10.17762/ijritcc.v11i2.6108
***Article Received**: 10 December 2022 **Revised**: 15 January 2023 **Accepted**: 27 January 2023*

Table 1. Models' performance

Model	Accuracy	Recall	Precision	F1-Score
LR	0.84	0.86	0.86	0.83
DT	0.78	0.70	0.83	0.69
Random Forest	0.88	0.86	0.86	0.83
LSTM	0.93	0.89	0.91	0.90

Table 2. Measures of metrics

Name of the metric	Measure of the metric
Accuracy	***(TP+TN)/((FP+FN+TP+TN))***
Precision	***TP/((FP+TP))***
Recall	***TP/((FN+TP))***

As a result, the letters TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and consequently False Negative.

Here, determine which of the four predictive models is best at predicting the level of student engagement, and which student activities are more important after testing the four models. increase. The results shown in Table 1 demonstrate that Student involvement levels can be predicted using machine learning algorithms.

The search for all algorithms is the same as shown in Table 1, but the LSTM algorithm outperforms the other three models in accuracy, F-1 score, accuracy, and precision. The majority of researchers predicting student performance [18] support these conclusions. Table 2 displays the metrics for the aforementioned metrics.

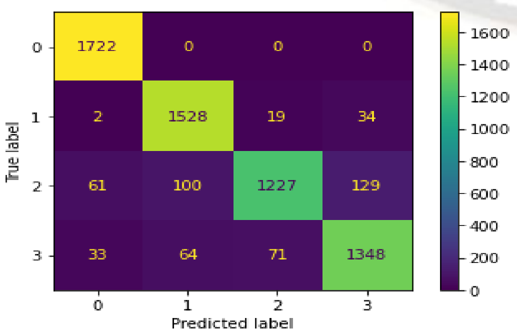

Fig 1: Confusion matrix for the Linear regression prediction model

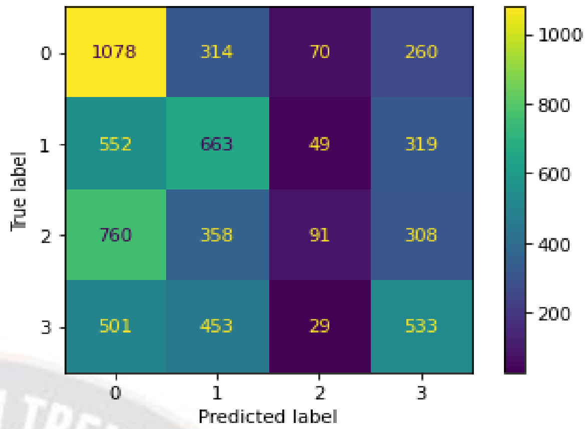

Fig 2: Confusion matrix for decision tree prediction model

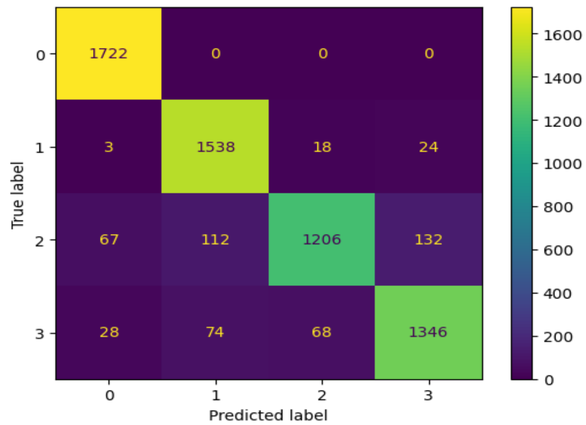

Fig 3: Confusion matrix for Random forest prediction model

The confusion matrices of the LR, DT, and Random Forest Classifier are depicted in the above figures 1,2, and 3. The algorithm with the highest measurements of Recall, Precision, and F-1 Score is Random Forest, as seen in the above figure.

Fig 4. Compares the four models' accuracy and displays the results. The most accurate algorithm was the LSTM and is the most appropriate algorithm for this study.

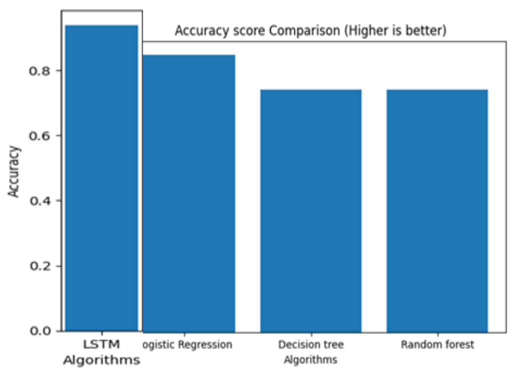

Fig 4. Accuracy of all models

As shown in Table 3, the Decision Trees and Random Forest algorithms are less accurate, but they are easier to understand the key characteristics of the data.

Table 3: Item-wise analysis of student performance

S. No	Salient Features	Importance
1	Total Logins	0310349
2	Activity inside content area	0.249501
3	Nbr. clicks	0.137109
4	Join session	0.100937
5	User Activity group	0.074397
6	Total items	0.034525
7	Time spent	0.031743
8	Time Spent session attendance	0.016438

V. Conclusion and Future Scope

This study's main focus has been on predicting student involvement using the information that students learned during an online session. Among the four models in this investigation, the LSTM algorithm has the highest accuracy. However, the DT and Random Forest algorithms are employed to identify the key data characteristics associated with the engagement elements. The overall percentile of test achievement, the amount of time spent on each session, and the knowledge levels of individual students are used to determine student involvement. The suggested system can help educational institutions when their ability to provide instruction is disrupted, whether it's because of a COVID-19-like pandemic or another form of natural disaster that forces them to close for a while.

Future studies will examine aspects of time associated with student assessment, mentor presence, and the effectiveness of online sessions.

References

[1]. M. H. Baturay, "An Overview of the World of MOOCs," *Procedia Soc Behav Sci*, vol. 174, pp. 427–433, Feb. 2015, https://doi.org/10.1016/j.sbspro.2015.01.685

[2]. I. Maiz, "Research on MOOCs: Trends and Methodologies."

[3]. Robert M. Carini and George D. Kuh, "Student engagement and student learning: Testing the linkages," *Res High Educ*, vol. 47, no. 1, Feb. 2006.

[4]. M. D. Dixson, "Measuring Student Engagement in the Online Course: The Online Student Engagement Scale (OSE)."

[5]. J. S. Lee, "The relationship between student engagement and academic performance: Is it a myth or reality?" *Journal of Educational Research*, vol. 107, no. 3, pp. 177–185, May 2014, https://doi.org/10.1080/00220671.2013.807491.

[6]. J. VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data."

[7]. M. Atherton, M. Shah, J. Vazquez, Z. Griffiths, B. Jackson, and C. Burgess, "Using learning analytics to assess student engagement and academic outcomes in open access enabling programs," *Open Learning*, vol. 32, no. 2, pp. 119–136, May 2017, https://doi.org/10.1080/02680513.2017.1309646

[8]. Mutahi, "Studying Engagement and Performance with Learning Technology in an African Classroom," *Association for Computing Machinery*, pp. 148–152, Mar. 2017.

[9]. Timothy Rodgers, "Student engagement in the e-learning process and the impact on their grades," *International Journal of Cyber Society and Education*, vol. 1, no. 2, Dec. 2008, Accessed: Jan. 18, 2023. [Online]. https://www.learntechlib.org/p/209167

[10]. Zhuang, "Understanding Engagement through Search Behaviour," *Association for Computing Machinery*, pp. 1957–1966, Nov. 2017, https://doi.org/10.1145/3132847.3132978

[11]. Arvid Kappas, "The Affective Computing Approach to Affect Measurement," vol. 10, no. 2, Sep. 2017, Accessed: Jan. 18, 2023. [Online]. https://doi.org/10.1177/1754073917696583

[12]. F. H. Veiga, "Proposal to the PISA of a New Scale of Students' Engagement in School," *Procedia Soc Behav Sci*, vol. 46, pp. 1224–1231, 2012, https://doi.org/10.1016/j.sbspro.2012.05.279

[13]. B. J. Mandernach, "Assessment of Student Engagement in Higher Education: A Synthesis of Literature and Assessment Tools," 2015.

[14]. B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel, "The validity and utility of activity logs as a measure of student engagement," in *ACM International Conference Proceeding Series*, Mar. 2019, pp. 300–309. https://doi.org/10.1145/3303772.3303789

[15]. R.-M. Conrad and • J Ana Donaldson, "Engaging the Online Learner: Activities and Resources for Creative Instruction."

[16]. C. Iwendi, S. Ponnan, R. Munirathinam, K. Srinivasan, and C. Y. Chang, "An efficient and unique TF/IDF algorithmic model-based data analysis for handling applications with big data streaming," *Electronics (Switzerland)*, vol. 8, no. 11, Nov. 2019, https://doi.org/10.3390/electronics8111331

[17]. C. Iwendi, S. Khan, J. H. Anajemba, M. Mittal, M. Alenezi, and M. Alazab, "The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems," *Sensors (Switzerland)*, vol. 20, no. 9, May 2020, https://doi.org/10.3390/s20092559

[18]. F. Ofori and R. Gitonga, "Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature-Based Review Cloud security View project Cloud security View project." [Online]. https://www.researchgate.net/publication/340209478