Activity Management System for Enhancing Online Classroom Interaction Using Support Vector Machine and Random Forest

Akaraj Chunhapinyokul*, Lachana Ramingwong, Sakgasit Ramingwong, Kenneth Cosh, and Narissara Eiamkanitchat

Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
Email: akaraj_c@cmu.ac.th (A.C.); lachana_r@cmu.ac.th (L.R.); sakgasit_r@cmu.ac.th (S.R.); kenneth_c@cmu.ac.th (K.C.); narissara_e@cmu.ac.th (N.E.)
*Corresponding author

Manuscript received June 9, 2025; accepted September 9, 2025; published February 11, 2026.

***Abstract*—This paper proposes an Activity Management System (AMS) that uses Support Vector Machine (SVM) and Random Forest (RF) algorithms to classify students into different interaction levels in online classrooms. The system captures behavioral features such as time spent online and chat activity to form data-driven student groups. The approach addresses the lack of interaction in virtual learning environments and promotes more equitable and engaging peer collaboration. The system's effectiveness was validated using real and CTGAN-generated data, showing improvements in interaction compared to random grouping.**

***Keywords*—online learning, student engagement, Support Vector Machine (SVM), Random Forest (RF), machine learning in education, interaction classification**

I. Introduction

Online education, especially post-pandemic, presents challenges such as student isolation and low interaction. Previous studies show that effective interaction strongly impacts learning outcomes. While traditional LMS platforms provide some interaction features, they often rely on random group formation. This paper introduces a machine learning-based AMS that dynamically groups students based on interaction data to enhance online peer engagement.

II. Literature Review

A. Student Interaction in Online Classrooms

Interaction is essential for learning and engagement, particularly in online environments where spontaneous communication is limited [1, 2]. During the pandemic, nearly 70% of students reported disengagement due to limited peer and instructor interaction [3]. While LMS platforms like Moodle and Canvas offer discussion forums and group assignments, these often rely on random group formation, which may not reflect students' behavioral engagement levels [4, 5]. Recent approaches have applied Machine Learning (ML) and behavioral analytics—such as login frequency and forum activity—to form groups more intelligently [6, 7].

B. Application of Support Vector Machine (SVM)

SVM is widely used to classify student engagement from LMS data due to its effectiveness with small and imbalanced datasets [8]. For instance, Jayaprakash *et al.* [9] used SVM to identify at-risk students based on forum and assignment activity, achieving high predictive performance. Compared to K-Nearest Neighbors (KNN), Naïve Bayes, and Neural Networks, SVM has shown superior results in engagement classification [10].

C. Use of Random Forest for Student Grouping

Random Forest (RF) is an ensemble method effective for predicting academic success and engagement [10, 11]. It reduces overfitting and captures complex patterns, making it suitable for behavior-based grouping.

D. Synthetic Data Generation Using CTGAN

CTGAN is designed for generating synthetic tabular data and is useful in educational settings with limited datasets. Xu *et al.* [12] developed CTGAN to synthesize interaction logs while preserving privacy, and Zhang *et al.* [13] reported increased model accuracy when using CTGAN-generated data for training.

III. Methodology

Research Objective:

The objective of this study is to investigate the effect of group formation based on intelligent techniques in enhancing online classroom interaction through an Activity Management System (AMS). The system supports intelligent group formation using decision tree-based algorithms and analyzes interaction logs collected during student use of the platform.

Hypotheses:

H1: Students who use the AMS during online learning will demonstrate significantly higher levels of peer interaction than those using other online learning support tools that allow activity creation but lack intelligent grouping mechanisms.

H2: Students grouped using machine learning algorithms (SVM and Random Forest) will exhibit significantly higher levels of interaction than those grouped randomly.

Expected Contributions:

The findings from this study can inform the development or enhancement of interaction-support tools for online education. Furthermore, they may assist instructors in making data-informed decisions to improve peer collaboration, engagement, and the overall online learning experience.

A. Development of the AMS Website

To address the research objective of enhancing student interaction in online classrooms through data-driven group formation, the Activity Management System (AMS) was developed. This system was designed in alignment with the study's research questions, particularly focusing on whether machine learning-based grouping methods improve interaction among students compared to random assignment. The AMS served as both a data collection platform and a grouping mechanism for experimental purposes.

The system was developed using Visual Studio Code as the integrated development environment. The backend was implemented using PHP and SQL, while Python was employed for machine learning operations. The frontend interface was developed using HTML and CSS to ensure user accessibility and responsiveness. A local database environment was established using XAMPP, which enabled secure and efficient storage of user interaction logs during the study period.

B. Participant Demographics and Data Collection

To support model development and evaluation, two groups of participants were recruited: a training group for machine learning model development and an experimental group for system testing. All participants were undergraduate students at a public university in northern Thailand and had prior experience with online or hybrid learning environments.

1) Training group (n = 22)

This group included 22 students aged between 19 and 22, primarily enrolled in computer science, information technology, or education-related programs. Interaction data were automatically logged by the AMS system during both individual and group-based online activities over a two-week period. Six key features were extracted for modeling: time spent online during regular platform usage (Tn), time spent in collaborative activities (Tg), number of messages during group learning (Mg), total message length in group activities (Lg), number of messages during individual learning (Mn), and message length in individual learning (Ln). These features represent both behavioral frequency and depth of engagement. A normalized weighted scoring approach was used to calculate overall interaction scores for classification.

2) Experimental group (n = 30)

This group comprised 30 student volunteers aged 19 to 24, representing diverse academic fields including humanities, social sciences, and STEM. Prior to group formation, a survey was conducted to assess their familiarity with each other: 63% reported knowing none of their peers, 25.9% knew 1–2 peers, and 11.1% knew 4–6 peers. These insights were used to contextualize the interaction outcomes observed during the experimental phase.

Demographically, both groups reflect typical regional university students and provide a reasonable basis for evaluating the model's performance and generalizability across varied online learning contexts.

C. Weighted Interaction Score Calculation

To classify student interaction levels, this study employed a weighted scoring formula that combines normalized behavioral features. Each feature was normalized using min-max normalization and multiplied by a corresponding weight that reflects its relative importance in measuring student engagement.

The interaction score (Score) was computed as follows:

$$\text{Score} = 0.35\cdot\text{Norm}(Tn)+0.35\cdot\text{Norm}(Tg)+0.10\cdot\text{Norm}(Mg)+0.10\cdot\text{Norm}(Lg)+0.05\cdot\text{Norm}(Mn)+0.05\cdot\text{Norm}(Ln) \quad (1)$$

Norm() refers to Min-Max normalization applied to each feature to scale values between 0 and 1.

D. Weight Derivation

The weights used in the interaction scoring formula—0.35 for online time, 0.10 for chat message count, and 0.05 for total message length—were determined through a hybrid approach that integrated Exploratory Data Analysis (EDA) with expert-informed heuristics. This methodological choice reflects a balance between empirical pattern discovery and theoretical alignment with pedagogical principles.

To strengthen construct validity, two domain experts specialized in learner engagement assessment and the application of machine learning in education, each with more than a decade of experience were involved in the weighting process.

The final weighting schema was selected based on a consensus that prioritized behavioral indicators most closely aligned with the interaction constructs defined in the study. This process ensures that the scoring mechanism is not merely data-dependent but also anchored in established pedagogical theory.

- Time-based features (online time) showed the strongest correlation with overall engagement and were therefore assigned the highest weight (0.35 each).
- Chat features from gamified sessions showed moderate influence on interaction and received medium weight (0.10 each).
- Chat features from normal sessions were found to be less frequent and less correlated with perceived engagement, thus assigned a lower weight (0.05 each).

The weights were normalized to ensure that their total equals 1.0, maintaining proportional influence.

E. Interaction Level Classification

After computing the normalized weighted score for each student, interaction levels were classified into four categories based on score percentiles:

Level 1 – Very Low: 0.00 – 0.25

Level 2 – Low: 0.26 – 0.50

Level 3 – Medium: 0.51 – 0.75

Level 4 – High: 0.76 – 1.00

This classification enables educators to better understand students' engagement and provides a foundation for personalized interventions or activity planning.

The distribution of students across the four interaction levels is summarized in Table 1.

Table 1. Classification results of student interaction levels

Interaction Level	Description	No. of Students
Level 1	Very Low	5
Level 2	Low	4
Level 3	Medium	7
Level 4	High	6

F. Synthetic Data Generation Using CTGAN

Due to the limited size of the training dataset (n = 22), this study employed a Conditional Tabular Generative Adversarial Network (CTGAN) to generate synthetic interaction data. CTGAN is designed to synthesize realistic tabular datasets, making it well-suited for educational contexts characterized by small sample sizes and class imbalance.

The model was trained on six behavioral features and produced 2,000 synthetic records, evenly distributed across four predefined interaction levels. This class-balanced dataset was combined with the original data to improve model robustness. Results indicated that the inclusion of synthetic data enhanced classification accuracy and generalization, particularly in identifying underrepresented interaction categories, while maintaining the statistical integrity of the original dataset.

G. Model Robustness and Generalization Potential

Although the dataset consisted of only 22 real samples, the use of Support Vector Machine (SVM) and Random Forest (RF) was justified based on their well-established effectiveness in small to medium datasets. Both algorithms are known for their resilience to overfitting—SVM through margin maximization and RF through ensemble learning. The input features, selected for their behavioral relevance (e.g., time spent online and message activity), were well-structured and contributed to meaningful pattern discovery despite limited data.

To validate model performance, a 5-fold cross-validation strategy was applied. This method helped assess the model's generalization by evaluating performance across multiple training-testing splits. The results indicated stable classification accuracy and confirmed that the models could generalize within the current dataset structure. However, due to the narrow demographic scope, caution is warranted in applying these findings to broader or more diverse populations.

Alternative models such as Logistic Regression and Decision Trees were considered but yielded lower accuracy and showed signs of overfitting or underfitting, especially with class imbalance. By contrast, SVM and RF offered better performance and robustness, justifying their selection for classifying students into interaction-level groups.

H. Model Testing

To evaluate the effectiveness of the proposed models, a final test was conducted with a group of 30 student volunteers. Before the grouping experiment, a pre-test questionnaire was administered to assess the participants' familiarity with one another. This step was taken to observe whether existing relationships among students might influence the dynamics within the experimental group activities, and to minimize potential bias in interpreting the outcomes.

The questionnaire revealed the following levels of prior acquaintance among participants: 63% of the students reported that they did not know any of the other participants. 25.9% indicated that they knew 1–2 people in the group. 11.1% stated they were familiar with 4–6 people.

This background information helped provide context for analyzing group behavior, particularly in assessing whether group cohesion or performance could be attributed to pre-existing relationships rather than the grouping method.

After collecting the questionnaire data, students were divided into groups using three different approaches:

- Support Vector Machine (SVM)-based grouping. Students were clustered based on their interaction levels predicted by the SVM model trained on the augmented dataset.
- Random Forest (RF)-based grouping. Similar to the SVM approach, but using the predictions made by the Random Forest model.
- Random grouping. Students were grouped randomly without considering interaction levels, serving as a baseline for comparison.

The performance and collaboration within each group were later evaluated to determine the effectiveness of each grouping strategy in promoting student interaction and engagement. This evaluation was crucial in verifying whether machine learning-based grouping provided meaningful improvements over random assignment.

I. Interaction Evaluation

To assess the effectiveness of each grouping method on learner-to-learner interaction, participants completed a questionnaire adapted from the study "Interactions Quality in Moodle as Perceived by Learners and Its Relation with Some Variables" by Ahmed Yousif Abdelraheem (*TOJDE*, vol. 13, no. 3, 2012). The questionnaire comprised 8 items measuring the quality of interaction using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

Each group of students was formed using one of three grouping methods:

- Random Grouping (NR)
- Support Vector Machine (SVM)-based Grouping
- Random Forest (RF)-based Grouping

In addition, scores from the reference study (REF_CP), which used Moodle as the platform, were used as a benchmark for comparison.

The questionnaire used to evaluate learner-to-learner interaction is presented in Table 2.

Table 2. Learner-to-Learner interaction questionnaire items (adapted from Abdelraheem, 2012)

No.	Question
1	The AMS system encouraged me to seek out additional resources from my classmates.
2	The AMS system was useful and effective.
3	The AMS system encouraged me to evaluate my learning in a good way.
4	The AMS system allows me to interact with my friends without it having to do with the lesson.
5	The AMS system helped me understand my ideas from a new perspective.
6	I interact with other students in the AMS system.
7	I am not ashamed to send messages and offer my ideas.
8	I have a good time for discussion among ourselves.

IV. Discussion and Result Analysis

To evaluate the impact of group formation methods on student interaction, responses from the post-activity questionnaire were analyzed across three grouping conditions: Moodle-based benchmark (REF_CP), Random Grouping (NR), and machine learning-based grouping—Support Vector Machine (SVM) and Random Forest (RF). The analysis proceeded as follows:

A. Descriptive Statistics

For each of the eight interaction-related questions, the mean and standard deviation were calculated separately for each grouping method. This approach enabled a clear comparison of learner-to-learner interaction levels across groups. Summary statistics showed that both SVM and RF groups consistently achieved higher interaction scores compared to random grouping.

B. Comparison against Benchmark

The results from the machine learning-based groups (SVM and RF) were compared with the control group (NR) and the Moodle-based benchmark group (REF_CP).

Both SVM and RF groups demonstrated higher average interaction scores than the random group (NR).

The SVM-based grouping achieved the highest scores across most questionnaire items, closely followed by the RF group.

These results indicate that using student behavior logs to inform group formation fosters greater collaboration and engagement compared to random assignment strategies traditionally employed in online classrooms.

C. Interpretation of Key Items

Notably, specific questionnaire items related to group communication, collaboration, and mutual understanding received particularly high scores in the SVM and RF groups. For instance, items such as:

"I interact with other students in the AMS system."

"I am not ashamed to send messages and offer my ideas." showed marked improvements compared to the random grouping condition.

This suggests that machine learning-based grouping contributed not only to more frequent interaction but also to the creation of psychologically safe environments, where students felt more comfortable engaging with their peers. These outcomes align with the study's main objective: enhancing student interaction in online classrooms through intelligent, data-driven group formation.

The results are illustrated in Fig. 1, which shows the mean Likert scores with standard deviation across all questionnaire items for each grouping method.

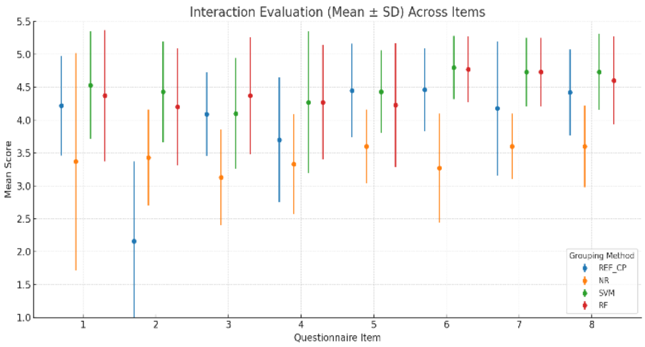

Fig. 1. Mean Likert scores (1–5) with standard deviation for each questionnaire item (Q1–Q8). The chart compares interaction levels across four grouping methods: REF_CP (Moodle-based benchmark), NR (Random Grouping), SVM (Support Vector Machine), and RF (Random Forest).

X-axis: Questionnaire Items (Q1 to Q8)
Y-axis: Mean Interaction Score (Likert Scale: 1 = Strongly Disagree to 5 = Strongly Agree)
Error bars: Represent ±1 standard deviation from the mean for each group

V. Conclusion

This study demonstrated that the proposed Activity Management System (AMS), which integrates machine learning techniques, can effectively enhance student interaction in online classrooms. By employing Support Vector Machine (SVM) and Random Forest (RF) algorithms to group students based on their behavioral data—such as online activity time and communication patterns—the system facilitated more productive peer interactions and collaboration compared to traditional random grouping methods.

Post-activity questionnaire results revealed that students grouped using machine learning, particularly via SVM, consistently exhibited higher levels of interaction than those assigned randomly. This underscores the potential of data-driven grouping to foster meaningful engagement in virtual learning settings. The interaction scoring model, derived from behavioral indicators such as time spent online and messaging activity, enabled nuanced differentiation of student participation levels without requiring complex or invasive metrics.

While overall outcomes were positive, one questionnaire item—related to gaining new perspectives—received a slightly lower average score compared to a prior benchmark using the Moodle platform. This suggests that while AMS effectively supports behavioral interaction, further enhancement of the system's cognitive engagement capabilities may be necessary.

Beyond student grouping, AMS also serves as a practical support system for instructors. The ability to classify interaction levels in real-time offers valuable insights that can inform instructional strategies, enabling timely interventions and reducing student isolation. This function aligns closely with the goal of increasing engagement and improving learning outcomes.

Finally, the conceptual framework presented in this study has practical applicability in broader educational contexts. It can be embedded into existing Learning Management Systems (LMS) as an intelligent alternative to manual or random group formation. By offering a behavior-driven grouping option, LMS platforms can better support personalized, interaction-centered learning experiences in online and hybrid environments.

Ethical Consideration

This study was approved by the Human Research Ethics Committee, Faculty of Medicine, Chiang Mai University (COA No. 010/67, CMUREC No. 66/344). All participants provided informed consent.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Akaraj Chunhapinyokul conducted the research design, system development, machine learning model implementation, data analysis, and manuscript preparation; Lachana Ramingwong provided guidance on methodology, supervised the research framework, and contributed to manuscript refinement; Sakgasit Ramingwong contributed to the conceptual design, provided technical insights on system architecture, and reviewed the manuscript; Kenneth Cosh supported with academic writing improvements, data interpretation, and critical review of results; Narissara

Eiamkanitchat assisted in data collection, system testing, and the preparation of research materials; all authors had reviewed and approved the final version of the manuscript.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Informatics, Mahasarakham University, the Faculty of Medicine, Chiang Mai University, and the Faculty of Engineering, Department of Computer Engineering, Chiang Mai University, for their administrative support during the research process.

REFERENCES

[1] M. G. Moore, *Theory of Transactional Distance*, Routledge, 1993.
[2] J. C. Richardson and K. Swan, "Examining social presence in online courses," *J. Asynchronous Learn. Netw.*, vol. 7, no. 1, pp. 68–88, 2003.
[3] B. Means and J. Neisler, "Suddenly online: A national survey of undergraduates during the COVID-19 pandemic," *Digital Promise*, 2020.
[4] I. Almarashdeh, "Sharing instructors' experience of LMS satisfaction in distance learning," *Comput. Human Behav.*, vol. 63, pp. 249–255, 2016.
[5] N. Capdeferro and M. Romero, "Are online learners frustrated with collaborative learning?" *Int. Rev. Res. Open Distrib. Learn.*, vol. 13, no. 2, pp. 26–44, 2012.
[6] D. T. Tempelaar, B. Rienties, and B. Giesbers, "Learning analytics in a data-rich context," *Comput. Human Behav.*, vol. 47, pp. 157–167, 2015.
[7] J. W. You, "Predicting course achievement using LMS data," *Internet High. Educ.*, vol. 29, pp. 23–30, 2016.
[8] M. Kloft, *et al.*, "Predicting MOOC dropout using machine learning," in *Proc. EMNLP Workshop*, 2014, pp. 60–65.
[9] S. M. Jayaprakash, *et al.*, "Early alert of academically at-risk students," *J. Learn. Analytics*, vol. 1, no. 1, pp. 6–47, 2014.
[10] M. A. Al-Barrak and M. Al-Razgan, "Predicting GPA using decision trees," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 528–533, 2016.
[11] C. Romero, S. Ventura, and E. García, "Data mining in Moodle," *Comput. Educ.*, vol. 51, no. 1, pp. 368–384, 2008.
[12] L. Xu, *et al.*, "Modeling tabular data using conditional GAN," *NeurIPS*, vol. 32, 2019.
[13] W. Zhang, C. Yin, and D. Wang, "Enhancing student prediction with CTGAN," *Int. J. Emerging Tech. Learn.*, vol. 15, no. 12, pp. 74–86, 2020.