International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I12P219
Volume 71 Issue 12, 195-200, December 2023

Original Article

Hybrid Learning Model Analytics to Predict Learning Style Behavior Clusters of Post-COVID-19 Learners in Higher Education

Kanakarn Phanniphong[1], Wongpanya S. Nuankaew[2], Patchara Nasa-Ngium[3], Pratya Nuankaew[4]

[1]*Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-ok, Bangkok, Thailand*
[2,4]*School of Information and Communication Technology, University of Phayao, Phayao, Thailand.*
[3]*Faculty of Science and Technology, Raja Maha Sarakham University, Maha Sarakham, Thailand.*

[4]*Corresponding Author : pratya.nu@up.ac.th*

Received: 21 August 2023 Revised: 21 November 2023 Accepted: 24 November 2023 Published: 06 December 2023

Abstract *- The impact of COVID-19 has forced the learning process to be organized into a hybrid learning environment that requires online and onsite learning services. Therefore, this research has three main objectives: 1) to cluster learners corresponding to learners' learning behaviors with a hybrid learning model, 2) to generate a predictive model for learner learning behavior clusters, and 3) to test the effectiveness of a predictive model for learning behavior clusters. The population and research sample consisted of 24 students enrolled in the course 221203[1] Technology for Business Application at the School of Information and Communication Technology, University of Phayao, in the second semester of the 2022 academic year. Research tools using supervised and unsupervised machine learning include K-Means, Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN), Neural Networks, Generalized Linear Model, and Support Vector Machine. The cross-validation approach and confusion matrix techniques were used to test the model with four metrics: Accuracy, Precision, Recall, and F1-Score. The research results showed that the learner clustering was appropriate with three clusters and consistent with the learning achievement of the learners. In addition, it was found that the cluster prediction model had a very high level of accuracy, with an accuracy of 96.67% and an S.D. of ±10.54%. Therefore, disseminating research findings in the public interest is appropriate.*

Keywords *- Educational data mining, Hybrid learning model, Learning analytics, Learning styles, Post-COVID-19.*

1. Introduction

The emergence of COVID-19 has dramatically changed how the educational process is organized. Face-to-face learning in public spaces and classrooms is not possible, and the effect causes learners and instructors to adjust [1, 2]. Like other countries, Thailand has also been affected by COVID-19 [3, 4]. Part of the significant impact is the lack of learning materials for rural people. Many students in Thailand have to study on phones with small screens. Moreover, some areas do not have a signal or network to enter the distance learning system [5–7].

However, the familiarity of students who have received online learning services from the network resulted in many students wanting to continue learning online. Therefore, educators and educational technologists developed a parallel learning process called "Hybrid Learning" [8–10]. Typically, hybrid learning allows learners to choose their learning styles. Teachers act as facilitators, providing learning resources and motivating them to learn. In addition, teachers organize two learning environments to arouse learners' desire to participate in activities comprising regular classroom learning and a virtual system. Both of these systems are operated by the teacher in parallel so that students can interact between the two systems [11].

From the source of the problem and the importance of changing student behaviour, researchers create learning strategies by incorporating hybrid learning through active learning to meet the needs of the learners. It consists of three main objectives. The first objective is constructing student clusters corresponding to student learning behaviours post-COVID-19, operating a hybrid learning style. The critical point of this purpose is to study the cluster-specific characteristics of the learners in each cluster. The second objective is to generate a predictive model for learner learning behaviour clusters. The goal of the second objective is to be able to adapt to future teaching and learning management. The last objective is to test the effectiveness of a predictive model for learning behaviour clusters. The goal is to select the most efficient model.

Population Sample and Research: the researchers controlled the learning activities in the hybrid model during the academic year 2022 in the second semester in the course 221203[1] Technology for Business Application at the School of Information and Communication Technology, University of Phayao. The research sample was a voluntary collaboration of 24 students who enrolled and received grades. The research tool uses supervised and unsupervised machine learning techniques, including K-Means, Decision Tree, Naïve Bayes, K-Nearest Neighbours (KNN), Neural Networks, Generalized Linear Model, and Support Vector Machine. The cross-validation approach and confusion matrix techniques were used to test the model with four metrics: Accuracy, Precision, Recall, and F1-Score. Using these tools and techniques is consistent with many research papers. Several research studies on the impact of COVID-19 on the educational quality improvement process encourage and drive researchers to have an interest in and need to develop a hybrid learning model to address current problems. The researchers are very hopeful that this research will steer the development of modern learning styles in the post-pandemic period of COVID-19 in line with learners' changing characteristics and behaviours and hope that this research will be helpful to the public.

2. Materials and Methods

2.1. Population and Sample

The research population was students enrolled in the course 221203[1] Technology for Business Application in Business at the School of Information and Communication Technology, the University of Phayao, during the second semester of the academic year 2022. There are 24 students, and the sample selection is a method of purposive sampling with the consent of the students in the course.

The learning activities were the hybrid learning style with online and onsite channels throughout the semester due to the post-COVID-19 situation. According to the seven chapters, there were fourteen activities: pre-test and post-test.

Chapter 1: Basic concepts of applying technology in business,
Chapter 2: Enterprise and business information system,
Chapter 3: Technology application for various businesses,
Chapter 4: Management information system,
Chapter 5: Business analytics, application, and software for business,
Chapter 6: Web technology and Innovation, and
Chapter 7: Applies technology in other areas.

2.2. Data Collection

Data were collected on the fourteen activities and the academic achievement data of the students who gave consent from the course 221203[1] Technology for Business Application in Business, as summarized in Tables 1 and 2.

Table 1 presents scores and statistics for each activity designed from the seven lessons. It found that overall, learners had higher post-test scores (mean = 7.27) than pre-test scores (mean = 3.95). In addition, the learners seem to pay more attention to post-test activities, as observed from the average time spent doing the post-test (average time = 04:41) than the time spent on the pre-test (average time = 03:44).

Table 1. Data collection

Activities	Min	Max	Mode	Median	Mean	S.D.	Avg.Time
1st Pre-Test	2.00	8.00	5.00	5.00	5.05	1.50	03:31
2nd Pre-Test	0.00	7.00	3.00	4.00	4.17	1.74	04:09
3rd Pre-Test	0.00	9.00	3.00	3.00	3.71	1.92	03:35
4th Pre-Test	1.00	6.00	4.00	4.00	3.55	1.36	03:59
5th Pre-Test	0.00	8.00	3.00	3.00	3.36	1.92	03:27
6th Pre-Test	1.00	6.00	4.00	4.00	3.59	1.30	04:44
7th Pre-Test	1.00	8.00	3.00	4.00	4.25	1.94	02:49
Avg. Pre-Test	0.71	7.43	3.57	3.86	3.95	1.67	03:44
1st Post-Test	5.00	10	10.00	9.00	8.36	1.68	02:40
2nd Post-Test	2.00	10	9.00	8.00	7.38	2.45	04:24
3rd Post-Test	2.00	10	7.00	7.00	6.91	2.11	03:56
4th Post-Test	2.00	9.00	8.00	6.50	6.18	2.28	04:53
5th Post-Test	2.00	10	7.00	7.00	6.36	2.30	05:50
6th Post-Test	2.00	10	8.00	8.00	7.52	2.16	06:17
7th Post-Test	2.00	10	10.00	9.00	8.21	2.17	04:52
Avg. Post-Test	2.43	9.86	8.43	7.79	7.27	2.16	04:41

Table 2. Academic achievement

Grade	Fix – Rate	Student	Percentage
A	80.00 – 100.00	7	29.17
B+	75.00 – 79.99	1	4.17
B	70.00 – 74.99	5	20.83
C+	65.00 – 69.99	4	16.67
C	60.00 – 64.99	4	16.67
D+	55.00 – 59.99	1	4.17
D	50.00 – 54.99	2	8.33
F	0.00 – 49.99	0	0.00
Total		24	100.00

Table 2 shows the students' academic achievement. It found that most learners had the highest A grade, seven students, representing 29.17%. Moreover, all learners in the course passed the criteria, with the number of students receiving a grade F equal to 0. Data from these two sections will be used to analyse learner-learning behaviour clustering and create a predictive model for learner-behaviour clustering to develop learning strategies in the following semester.

2.3. Research Methodology and Tools

The methodological design and selection of research tools were divided into two sections. The first part is the development of the learner's learning cluster. The researchers used the K-Means clustering technique to cluster and applied the cluster selection method from the elbow technique. The k-means technique calculates the distance between data sets using Euclidean distance calculations [12]. Members in the same cluster will have a close distance from each other. The optimal K or cluster number was determined using the Davies-Bouldin Index (DBI) technique. DBI is a clustering quality criterion independent of the number of clusters to be analysed. In addition, DBI does not depend on the method of segmentation. Therefore, DBI was used as a criterion to compare the differences between cluster number determinations. DBI is calculated using the ratio between the sum of the distribution of the data in the group and the distance between the groups; with good segregation of groups, the distribution of groups is small, and the distance between groups must be significant.

The second part is to build a learner behaviour cluster prediction model. The researchers used six supervised learning techniques, including Decision Tree, Naïve Bayes, K-Nearest Neighbours (KNN), Neural Networks, Generalized Linear Model, and Support Vector Machine. The tools used in this section are techniques to predict the desired outcome of predicting future membership based on historical data. Each technique has different advantages and disadvantages, as follows.

Decision trees [13] are an easy-to-understand technique based on an inverted tree structure. It is, therefore, trendy to use in predictive modelling. However, the decision tree model has a problem when the model has a very high accuracy, known as overfitting, which model is not practical.

Naïve Bayes [14] is a technique that uses probabilities to make decisions and provides options for making predictions. The Naïve Bayes model's target will predict the occurrence of an event by increasing the possibility of an event occurring, which is often used to analyse data with a dependent event.

K-Nearest Neighbour (KNN) is a data classification method. It is classified as a supervised machine learning algorithm that already knowing the answer. Using KNN analyses, the new data from the original data are closest to each other by a given number K. It is explained that determining the value of K determines how many data are analysed that are closest to the data to be classified.

Neural Networks are the collection of algorithms and data that mimic how the human brain works and attempt to recognize the underlying relationships of the studied data to be used in developing a prototype. In this sense, artificial neural networks are the development of systems based on innate neurons. Therefore, scientists prefer to use Neural Networks to solve highly complex problems.

The Generalized Linear Model (GLM) is the model which extends from the General Linear Model (GLM) for creating the predictive models and linear relationship between class and covariates, which covers both continuous and discrete results based on the distribution of exponential family by random component and link function.

Support Vector Machine (SVM) is one of the machine learning models used to classify data by creating a hyperplane and finding the best line. Classification of data on a multidimensional uses an optimized selection section called feature selection, where the selection structure is derived from the data taught to the system. The number of frames used to describe a particular case is called a vector. Thus, the SVM model aims to distinguish a group of vectors.

After obtaining models from each technique, all models will be sent to select suitable models for deployment by model performance testing. The method of model performance testing is described in the next section.

2.4. Research Analysis and Interpretation

Results analysis and research interpretation clarified the most effective model selection process following the third research objective. The testing process is divided into two main parts: dividing the data to test the model and using four metrics to measure model performance.

Splitting the data to test the model uses a technique known as the cross-validation method. Cross-validation divides the data into equal partitions and defines some data, forming a model known as the training data set. After that, the rest of the data will be tested, called the testing data set. This test section used four metrics based on the Confusion Matrix methodology to measure model performance: Accuracy, Precision, Recall, and F1-Score.

Accuracy measures overall model performance calculated as the results of correct predictions divided by the total number of data sets. While precision is a performance of predictive abilities considered separately by class, it can be calculated from the functional accuracy of each class divided by the number of members of each class.

Recall is the predicted value corresponding to the actual value that appears in the class. It is calculated as the true predicted value divided by the amount of data present in the class. Finally, F1-Score is a composite measure that combines precision and recall to indicate class-by-class performance as calculated from Equation 1.

$$2*(precision * recall) / (precision + recall) \quad (1)$$

3. Research Results

The results of the study classified reporting into three essential areas: reporting appropriate learner clustering studies, developing suitable learner cluster prediction models from each technique, and selecting the most effective model.

3.1. Appropriate Learner Clustering

The techniques used for clustering and selecting optimal cluster numbers were K-Means and Davies-Bouldin Index (DBI) techniques. The results of clustering and selecting optimal cluster numbers are reported in Table 3.

Table 3 shows the performance measurement results from K-Means clustering using the BDI technique. The BDI with the most negligible dispersion, which means that the group members are well clustered, is K = 3 clusters.

Therefore, it can be concluded that the number of three clusters is suitable for developing a model for predicting groups of learners based on learning behaviour.

Table 3. Appropriate learner clustering

Number of Clusters	Number of Items	DBI
2	Cluster_0 = 6 items, Cluster_1 = 18 items	0.110
3	Cluster_0 = 4 items, Cluster_1 = 5 items, Cluster_2 = 15 items	0.094
4	Cluster_0 = 11 items, Cluster_1 = 8 items, Cluster_2 = 3 items, Cluster_3 = 2 items	0.110
5	Cluster_0 = 2 items, Cluster_1 = 7 items, Cluster_2 = 6 items, Cluster_3 = 3 items, Cluster_4 = 6 items	0.104

Table 4. Analytical results from six classifier techniques

Class	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Classifier	**Decision Tree**			**Naïve Bayes**		
Cluster_0	50.00%	25.00%	33.33%	75.00%	75.00%	75.00%
Cluster_1	100.00%	100.00%	100.00%	80.00%	80.00%	80.00%
Cluster_2	82.35%	93.33%	87.50%	86.67%	86.67%	86.67%
Accuracy	85.00%			81.67%		
S.D.	±19.95%			±33.75%		
Classifier	**K-Nearest Neighbors**			**Neural Networks**		
Cluster_0	100.00%	25.00%	40.00%	0.00%	0.00%	-
Cluster_1	75.00%	60.00%	66.67%	50.00%	60.00%	54.55%
Cluster_2	78.95%	100.00%	88.24%	88.24%	100.00%	93.75%
Accuracy	80.00%			73.33%		
S.D.	±35.83%			±33.52%		
Classifier	**Generalized Linear Model**			**Support Vector Machine**		
Cluster_0	100.00%	75.00%	85.71%	0.00%	0.00%	-
Cluster_1	100.00%	100.00%	100.00%	0.00%	0.00%	-
Cluster_2	93.75%	100.00%	96.77%	62.50%	100.00%	76.92%
Accuracy	96.67%			61.67%		
S.D.	±10.54%			±36.05%		

Table 5. The most effective model

	True Cluster_0	True Cluster_1	True Cluster_2	Precision	Recall	F1-Score
Pred. Cluster_0	3	0	0	100.00%	75%	85.7%
Pred. Cluster_1	0	5	0	100.00%	100%	100.0%
Pred. Cluster_2	1	0	15	93.75%	100%	96.8%

3.2. Learner Cluster Prediction Models

The predictive model development tools selected the best predictive classifier from six techniques: Decision Tree, Naïve Bayes, K-Nearest Neighbours (KNN), Neural Networks, Generalized Linear Model, and Support Vector Machine.

Table 4 shows the results of the analytical results of the six techniques of the predictive model analysis. It was found that the Generalized Linear Model developed model had the highest accuracy, with an accuracy of 96.67% and a well-distributed standard deviation with an S.D. of ±10.54%. It is reasonable to use this model in an application where the performance test model is presented in Table 5.

3.3. The Most Effective Model

This section presents the model performance certified as the best model for this research, a Generalized Linear Model development, as detailed in Table 5.

Table 5 summarizes the performance analysis of the model developed by the Generalized Linear Model technique. It found that the model's performance was accurate and accurate across all metrics.

4. Research Discussion

This discussion focuses on research objectives, which researchers can summarize into three key areas: student behaviour clustering, development of student behaviour cluster prediction models, and testing the model's ultimate efficacy.

For the results of the post-COVID-19 learner learning behaviours of university students in this study, researchers used K-Means to cluster and the Davies-Bouldin Index (DBI) to select the best optimal number of clusters. The researchers found that a K equal to three clusters was the most suitable, with the displayed DBI values showing that they were adequately coalesced, as summarized in Table 3. Therefore, the researchers used such members and clusters to develop a prediction model.

However, by dividing the cluster into three clusters, some groups had a more significant number of learners than others (Cluster_0 = 4 items, Cluster_1 = 5 items, Cluster_2 = 15 items). Researchers had to work hard to maximize the accuracy of their prediction models due to the relatively high heterogeneity of membership numbers.

In developing a prediction model with six supervised learning techniques, the researchers succeeded in developing models from different styles with different outcomes, as shown in Table 4. Primarily, researchers have found that some techniques achieve high predictive accuracy, but some caveats need attention. For example, when developing models using Neural Networks and Support Vector Machine techniques. However, it was found that the model has a relatively high predictability; there are some classes that the model cannot predict correctly, so researchers need to ban these two models. Moreover, when competing with the rest of the techniques, the researchers found that model development using the Generalized Linear Model technique had the highest accuracy, with an accuracy of 96.67% and an S.D. of ±10.54%. The researchers, therefore, selected this model as a prototype of a cluster prediction model of university student behaviour in the post-COVID-19 period. The researchers tested this model's effectiveness, as shown in Table 5.

Finally, researchers tested the performance of selected models modelled using the Generalized Linear Model techniques, as shown in Table 5. Table 5 shows that all classes can perform efficiently, i.e., the precision of Cluster _0 and Cluster_1 equals 100%, the recall of Cluster_1 and Cluster_2 equals 100%, and the F1-Score of Cluster_1 equals 100%.

Through the report and discussion of the findings, the researchers can conclude that the research achieved all its objectives and achieved all its intended goals.

5. Conclusion

This research examines the context of learners in the post-COVID-19 situation. The researchers aim to study and find suitable learning styles for learners, and this research has three crucial research goals. The first objective is to contain learner behaviour clustering in a hybrid learning management style. The second objective is to create a cluster prediction model for the behaviour of students who have managed hybrid learning in the post-COVID-19 period. The third objective is to test the performance of the selected model.

This research was accepted and participated in the research project by students from the School of Information and Communication Technology, University of Phayao, during the academic year 2022 in the second semester. The population and sample were 24 students enrolled in the course 221203[1] Technology for Business Application. The data used were records of the students' pre-test and post-test learning activities and grades of all students.

The tools used for this research were the implementation of supervised and unsupervised machine learning techniques, including K-Means, Decision Tree, Naïve Bayes, K-Nearest Neighbours (KNN), Neural Networks, Generalized Linear Model, and Support Vector Machine.

The research results achieved all objectives, and the researchers developed three learner-learning behaviour clusters that could be used to develop high-performance predictive models, with an accuracy of 96.67% and an S.D. of ±10.54%. The results of this research presentation of a hybrid learning method during the post-COVID-19 situation among students at the University of Phayao. The researchers hope this research will be a guideline for educators and researchers to apply in other research.

6. Research Limitations

Limitations of the research: The researchers found that most learners focused and participated in the pre-test activities at a low level. Part of the reason is that the teacher informs the students that the score in this section is used as a guide for only learning preparation; it is not calculated in grading, so the students pay attention to this activity at a low level. Researchers must develop strategies for future research to raise awareness of all learning activities.

Funding Statement

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao.

References

[1] Mohammod Moninoor Roshid, and Prodhan Mahbub Ibna Seraj, "Interrogating Higher Education's Responses to International Student Mobility in the Context of the COVID-19 Pandemic," *Heliyon*, vol. 9, no. 3, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ritika Mahajan et al., "COVID-19 and Management Education: From Pandemic to Endemic," *The International Journal of Management Education*, vol. 21, no. 2, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Attasit Srisubat et al., "Effectiveness of Favipiravir Monotherapy in the Treatment of COVID-19: Real World Data Analysis from Thailand," *The Lancet Regional Health - Southeast Asia*, vol. 11, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] N. Leelawat et al., "Twitter Data Sentiment Analysis of Tourism in Thailand During the COVID-19 Pandemic using Machine Learning," *Heliyon*, vol. 8, no. 10, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Catarina Félix de Oliveira et al., "How Does Learning Analytics Contribute to Prevent Students' Dropout in Higher Education: A Systematic Literature Review," *Big Data and Cognitive Computing*, vol. 5, no. 4, pp. 1-33, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Marzieh Karimi-Haghighi, Carlos Castillo, and Davinia Hernández-Leo, "A Causal Inference Study on the Effects of First Year Workload on the Dropout Rate of Undergraduates," *International Conference on Artificial Intelligence in Education*, pp. 15-27, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Joana R. Casanova et al., "Dimensionality and Reliability of a Screening Instrument for Students at Risk of Dropping Out from Higher Education," *Studies in Educational Evaluation*, vol. 68, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Saiful Amin et al., "The Effect of Problem-Based Hybrid Learning (PBHL) Models on Spatial Thinking Ability and Geography Learning Outcomes," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 19, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Natela Doghonadze, Tamari Dolidze, and Natia Vasadze, "Face-to-Face, Hybrid and Online English as a Foreign Language Learning Efficiency in Higher Education (Georgian and Italian students' views)," *Journal of Education in Black Sea Region*, vol. 7, no. 1, pp. 120-143, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Bambang Hariadi et al., "Hybrid Learning by Using Brilian Applications as One of the Learning Alternatives to Improve Learning Outcomes in College," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 10, pp. 1-10, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Pratya Nuankaew, Patchara Nasa-Ngium, and Wongpanya S. Nuankaew, "Self-Regulated Learning Styles in Hybrid Learning Using Educational Data Mining Analysis," *2022 26th International Computer Science and Engineering Conference (ICSEC)*, pp. 208-212, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Greg Hamerly, and Jonathan Drake, "Accelerating Lloyd's Algorithm for K-Means Clustering," *Partitional Clustering Algorithms*, pp. 41-48, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Raza Hasan et al., "Student Academic Performance Prediction by Using Decision Tree Algorithm," *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Jingnian Chen et al., "Feature Selection for Text Classification with Naïve Bayes," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5432-5435, 2009. [CrossRef] [Google Scholar] [Publisher Link]