Self-Regulated Learning Model in Educational Data Mining

https://doi.org/10.3991/ijet.v17i17.23623

Pratya Nuankaew(✉)
School of Information and Communication Technology, University of Phayao, Phayao, Thailand
pratya.nu@up.ac.th

Abstract—Artificial intelligence technology brings wide impacts on several dimensions. The impact on the education system is that educational technology has been disrupted, it radically changed the paradigm of learning management. Therefore, this research aimed to study the paradigm shift of the education system focusing on the deployment of artificial intelligence technology to support the learning model in the era affected by the COVID-19 pandemic. There are two research objectives: (1) to study an appropriate self-regulated learning model with data mining techniques for designing appropriate online learning management, and (2) to study the learning achievement factors of learners by applying blended learning and self-regulated learning techniques. The samples were 26 students at the University of Phayao who enrolled in the course 221203 Technology for Business Application in the 2nd semester of the academic year 2020. The research tool is a statistical analysis and machine learning tool. It consists of analyzing pre-test scores, post-test scores, midterm scores, final scores, academic achievement, clustering analysis, and clustering performance. As a result, it found that learners had five reasonable clusters for the academic achievement learning model. The results specified the different learning styles of the learners in two dimensions including online and offline scenarios. Therefore, in future work, the researcher looks forward to performing research in the scope of identifying the suitability and the necessity of converting the face-to-face learning model to a fully online learning model.

Keywords—blended learning model, eruptive technologies in learning, educational data mining, educational disruptive technologies, self-regulated learning model

1 Introduction

The transformation of today's technology engages awareness in the context of the digital world. There is a serious impact on the dimensions of the education system that affects all levels of education [1]–[3]. Along with educational technology and artificial intelligence technology, these technologies are incorporated to support the development of a new body of knowledge known as "Educational Data Mining: EDM" [4]–[7].

The conducted research included dimensions of behavioral, cognitive, and attitudes toward learning styles [6]. On the other hand, there was a researcher interested in the development of learning tools for monitoring forecasting [5], [7]. In addition, modern instructional technologies such as online learning, learning analytics, cross reality, machine learning, simulations, and online laboratories have become increasingly essential for all educational setting levels. This tendency has been reinforced by the growing digitalization, personalization, and internationalization of education at various dimensions. Moreover, the COVID-19 pandemic crisis has explicitly created a phenomenon of accelerating and strengthening these efforts. Although the COVID-19 pandemic will end in the future, many measures taken in the educational process will likely become part of the learning management approaches including common methods, tools, and technologies.

Consequently, this research aims to study the paradigm shift of the education system focusing on the deployment of artificial intelligence technology to explain the learning model during the COVID-19 pandemic situation. The learning model refers to the management in succeeding in the learning achievement. The research objectives are (1) to study an appropriate self-regulated learning model with data mining techniques for designing appropriate online learning management and (2) to study the learning achievement factors of learners by applying blended learning and self-regulated learning techniques. The researcher hypothesized the belief that the learners clustering according to the self-regulated learning behavior could promote higher learning achievement.

The scope of the research was 26 students, who enrolled in the course 221203, Technology for Business Application in the second semester of the academic year 2020 at the Department of Business Computer, School of Information and Communication Technology, the University of Phayao. The research process is a combination of two areas of knowledge: (1) the learning theory consists of blended learning and self-regulated learning, and (2) data mining technology in a new area of research, known as "educational data mining". The researcher selected the process of data mining methodology named "CRISP-DM: Cross-Industry Standard Process for Data Mining". It consists of six phases including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The definitions and implementation of the process are detailed in the literature reviews and methodological section.

2 Literature reviews and related works

2.1 Collaborative learning

Collaborative learning is a learning strategy, a structural design of organizing learning activities, and an environment to create learning achievement focused on group activities where group members learn to achieve their goals [8]–[12]. The elements and importance of the collaborative learning style are implemented in a wide variety of formats supported and mastered by instructors of different discipline backgrounds and teaching traditions. The learning elements are: (1) learning is an active and constructive process, (2) learners are diverse, (3) learning is inherently social, (4) learning depends on rich contexts, and (5) learning has affective and subjective dimension [8].

In addition, collective learning approaches are diverse and depend on the deployment process in different disciplines. Examples of collaborative learning management are (1) cooperative learning, (2) problem-centered instruction, (3) guided design, (4) discussion groups and seminars, (5) simulation techniques, (6) peer teaching, (7) cases design, (8) supplemental instruction, and (9) learning community activities [8]–[10]. These collaborative learning styles play a role in two main areas including the instructors and the learners, who participate in the activity together.

Therefore, the roles of instructors and learners in collaborative learning are indicated as the following functions. As the instructors aspect, they must be the ones who embrace the upcoming changes, either in activities that the learners may succeed or fail. Instructors need to be flexible and adaptable to various pressures. In addition, the instructor must perform four key roles including a facilitator, a mentor, an activity manager, and a learner assessor. The facilitator performs three functions providing services, providing comfort, and providing the student's needs. The mentor is responsible to provide some information to the learner and controlling learning activities so that the learner does not confuse with the knowledge point [13]. In the activity manager role, the instructor is responsible for planning group activities, planning the use of the activity duration, and controlling the production of work to be consistent with the activity goals. The final role of the instructor is to evaluate the understanding, knowledge, and perception of learners. On the other hand, learners play a five-key role: (1) creator of learning objectives, (2) planner for self-activities and solving problems, (3) organizers of knowledge, work pieces, and projects, (4) presenter of responsible activities, and (5) assessor for self-activities.

The collaborative role of the instructor and the learner leads to the process of creating knowledge. The tools and technologies used to collaborate in the development of knowledge for educational innovation are blended learning and self-regulated learning theories presented in the next section.

2.2 Blended learning and self-regulated learning

The concept of blended learning has been defined by academics, educators, and information technologists in several ways, such as hybrid, flexible, integrated, multimethod, and mixed mode learning [14]–[19]. In general, the definition of blended learning is a flexible learning model that aims to integrate both learner and instructor activities to manage activities through a variety of media and technology.

Educational technologists also define blended learning. It means conducting educational theory in combination with modern technology including online learning, distance learning, e-learning, mobile learning, digital learning model, and so on [20], [21]. However, technologies are not the only tool that supports the teaching and learning process. For instance, there is a tool that interferes with learning, known as "disruptive technologies in the education sector" [14], [22], [23]. These modern tools attracted the attention of learners at the same time. Examples of modern technology are Artificial Intelligence (AI), Augmented Reality (AR), Virtual Reality (VR), Chat-based Collaboration Platforms, Web-based services for online booking, Online Learning Courses Platforms, and Educational Robots for Internet-of-Things Supported Collaborative Learning (IoTSCL) technology [24]–[26]. It offered an enormous of available tools that support and impede the learning process. Therefore, choosing an appropriate educational theory for the transformation of technology is very imperative along with the behavior of learners who are always addicted to technology and lose their interest in learning [27]. The educational theory with a solution to this problem is the self-regulated learning theory.

Self-regulated learning (SRL) is the theory of social learning processes [21], [28]–[33]. It includes the cognitive, metacognitive, behavioral, motivational, and emotional affective aspects of learning. Educators influencing the theory of self-regulated learning consist of several researchers and educators. There are two very important theories including Bandura and Zimmerman Theory. Bandura offers a process for self-regulated learning in three components: (1) self-observation, (2) judgmental process, and (3) self-reaction. Bandura's theoretical framework is shown in Figure 1.

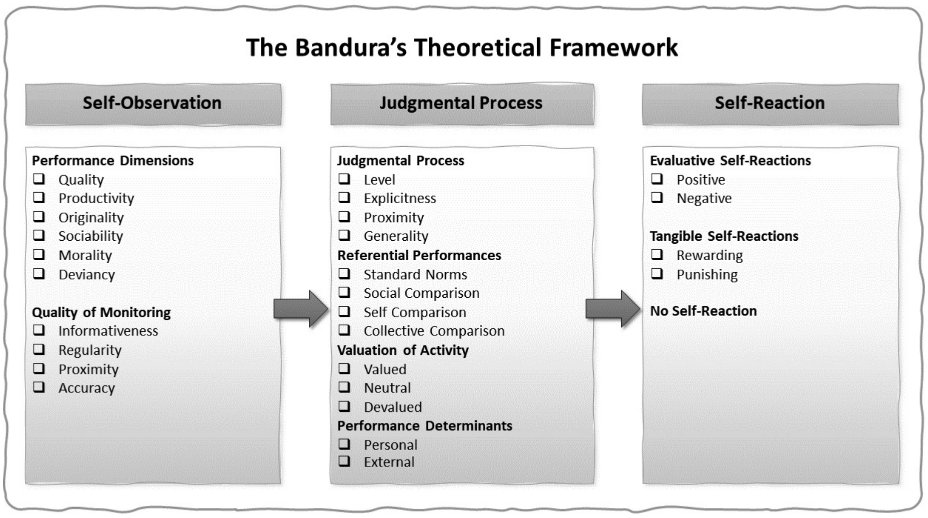

Fig. 1. The Bandura's theoretical framework [30]

According to Figure 1, Bandura's theoretical framework presents the process of the self-regulated learning model which consists of three main phases. The first phase is self-observation, which composes of two sub-process including (1) performance dimensions (quality, productivity, originality, sociability, morality, and deviancy dimension), and (2) quality of monitoring (informativeness, regularity, proximity, and accuracy quality). The second phase is the judgmental process, which consists of four sub-process including (1) personal standards (personal level, personal explicitness, personal proximity, and personal generality), (2) referential performances (standard norms, social comparison, self-comparison, and collective comparison), (3) valuation of activity (valued, neutral, and devalued), and (4) performance determinants (personal and external determinant). The last phase is the self-reaction, which contains three sub-process including (1) evaluative self-reactions (positive self-reaction and negative self-reaction), (2) tangible self-reactions (rewarding and punishing self-reaction), and (3) no self-reaction. The key tenet of Bandura's theory is being able to assess yourself according to your goals and reward yourself.

On the other hand, Zimmerman's theory aims to achieve learning with a social and cognitive perspective on self-regulated learning styles. His work covered social cognitive theory, such as individuals gaining knowledge by observing others and social interaction [34]. The model of self-regulated learning that the researcher attention to is a model of cooperation between Barry J. Zimmerman and Adam R. Moylan. It presented the concept of self-regulation as metacognition and motivation intersect [33], shown in Figure 2.

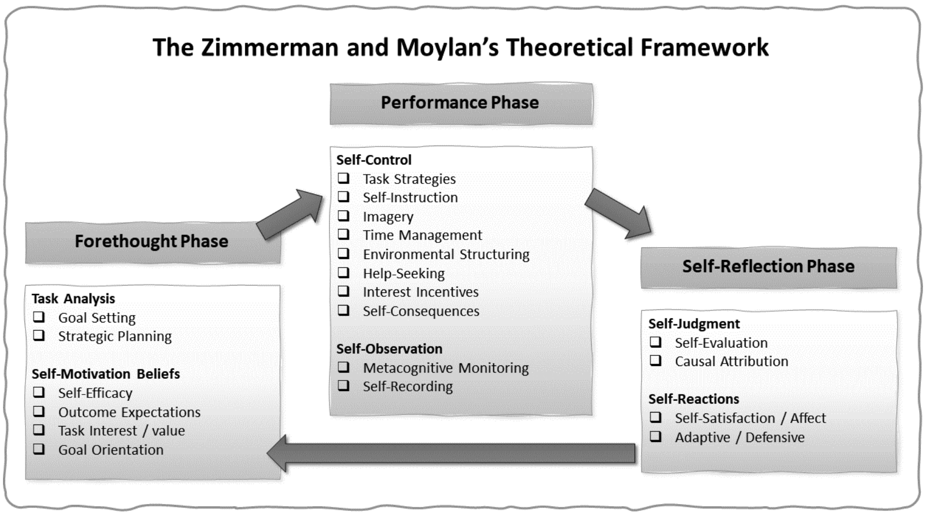

Fig. 2. The Zimmerman and Moylan's theoretical framework [33]

According to Figure 2, Zimmerman and Moylan's theoretical framework presents Zimmerman's Cyclical Phases Model of the self-regulated learning model which consists of three main phases [33] including the Forethought, Performance, and Self-reflection phase. In the forethought phase, students analyze work, define goals, and plan how to reach them. Moreover, several inspiring beliefs will motivate the process and influence the motivation of learning strategies. In the performance phase, students monitor their progress and employ multiple self-control strategies to keep themselves mentally. It engaged and motivated students to complete their tasks. Finally, in the self-reflection phase, students assess how they perform their work by identifying their successes or failures. These traits generate self-reactions that can be a positive or negative effect on the way students later approach work in acting.

As the study and review of the literature on blended learning [15]–[18], [35] and self-regulated learning [28]–[30], [33], [34], it found that both theories are very consistent with current changes in learner behavior, such as mobile addition, social media addiction, attention deficit hyperactivity disorder: ADHD [27], [36], [37], and so on. However, to make this research valuable and innovative, the researcher applied modern technology in data analysis using artificial intelligence technology to make it very diverse. This concept is known as "EDM: Educational Data Mining" discussed in the next section.

2.3 Educational data mining technology

Educational data mining technology refers to the processes of artificial intelligence and machine learning technology through techniques, tools, and research designed to automatically extract insights, knowledge, and patterns from large-scale data repositories generated by related activities. It aims to study and research the scope of educational institutions, learners' behavior, educational models, and academic achievement [5], [38].

In addition, the researcher defines the scope of research in educational data mining in five areas. The first area is the Academic Analytics (AA) and Institutional Analytics (IA), which involves compiling, analyzing, and visualizing the activity of the academic program. For example, course analysis, degree analysis, student fee income research, course assessment, resource allocation, and management of in-depth understanding of the institution. The second area is the Teaching Analytics (TA), which refers to the analysis of teaching activities and performance data including the design, development, and assessment of learning activities. The third area is Data-Driven Education (DDE) and Data-Driven Decision-Making in Education (DDDM). The systematic collection and analysis of different types of educational data guide a wide range of decision-making to improve learners' and institutions' success. The fourth area is Big Data in Education (BDE), which refers to the analysis of big data applied to data from the educational environment. The final area is Educational Data Science (EDS), which is defined as the use of collected data from the educational environment to set up educational problems [4].

Furthermore, the components of educational data mining are made up of three major domains: (1) Computer Science, (2) Statistics, and (3) Education. The conceptual framework of the educational data mine is shown in Figure 3.

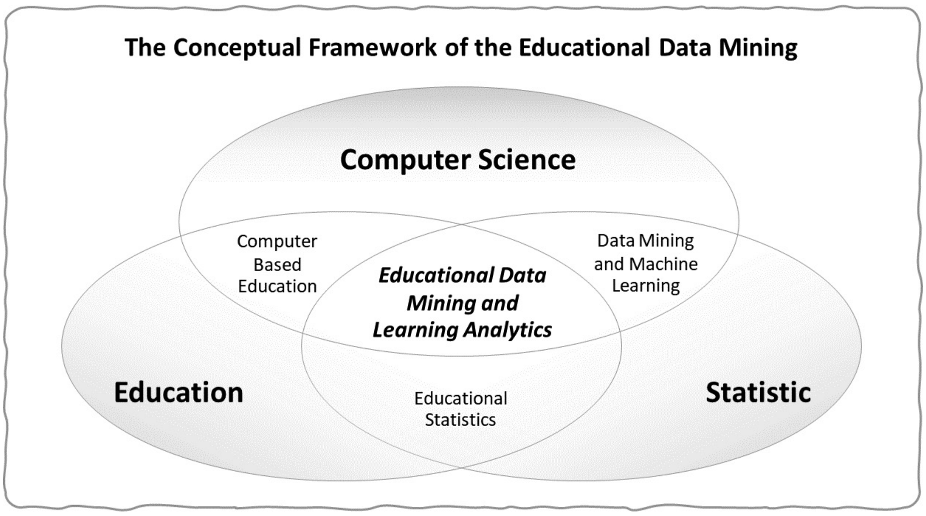

Fig. 3. The conceptual framework of the Educational Data Mining [4]

Figure 3 shows the composition of Educational Data Mining. It shows the area of overlap in three subsections: (1) Computer based education, (2) Educational statistics, and (3) data mining and machine learning. Combining all the elements are "Educational Data Mining (EDM) and Learning Analytics (LA)" [4]. According to studies and literature review, educational data mining applied a data mining tool, known as "CRISP-DM: Cross-Industry Standard Process for Data Mining" [39]. It is a tool for controlling research methods. The CRISP-DM model and phases are presented in Figure 4.

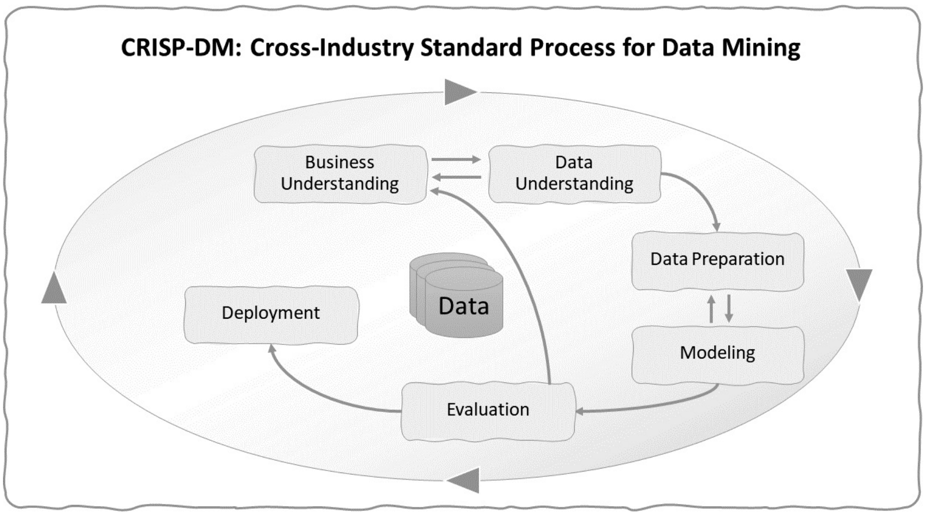

Fig. 4. CRISP-DM: Cross-Industry Standard Process for Data Mining [39]

Figure 4 shows the CRISP-DM: Cross-Industry Standard Process for Data Mining model that composes of six phases including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The details of the CRISP-DM process are described in the research methodology section.

3 Research methodology

As mentioned in the scope of the research and literature reviews, the entire research process consists of six phases. The steps and details can be demonstrated by the contents, shown in Figure 4.

3.1 Business understanding

The business understanding is the first phase in the CRISP-DM process. It focuses on understanding the research problems and converting them into research questions for analysis of data mining techniques and execution plans [39].

The purpose of this phase is to create and understand a research problem. This research aims to study the paradigm shift of the education system focusing on the deployment of artificial intelligence technology to explain the learning model during the COVID-19 pandemic situation.

The learning model refers to the management of success in learning achievement. Therefore, this research has two objectives: (1) to study an appropriate self-regulated learning model with data mining techniques for designing appropriate online learning management, and (2) to study the learning achievement factors of learners by applying blended learning and self-regulated learning techniques.

The researcher hypothesized the belief that clustering of learners according to the self-regulated learning behavior could promote higher learning achievement.

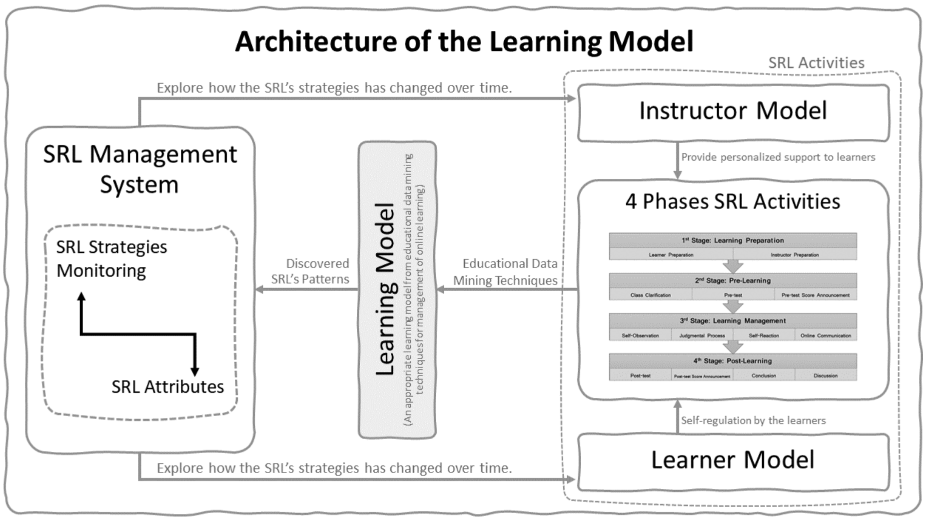

Fig. 5. The architecture of the learning model

Figure 5 shows the architecture of the learning model. There are three main sections in this architecture. The first section is the self-regulated learning activities discussed in the data understanding section. The second section is the learning model, which is explained in the modeling and evaluation section. Finally, the last section is the self-regulated learning (SRL) management system section, which explores the applied application in the deployment section.

3.2 Data understanding

The process of data understanding begins with gathering preliminary data and performing activities to familiarize with the common data. After that, the collected data will be reviewed to determine the correctness of the data and decide whether all data is used, or some data needs to be selected for analysis [39].

To recognize the data, the researcher determined the scope of the research by selecting 26 students who enrolled in the course 221203, Technology for Business

Application in the second semester of the academic year 2020. The learning method is the provision of online learning. The processes used during teaching and learning adopted the blended learning and self-regulated learning approaches in main activities, as shown in Figure 6.

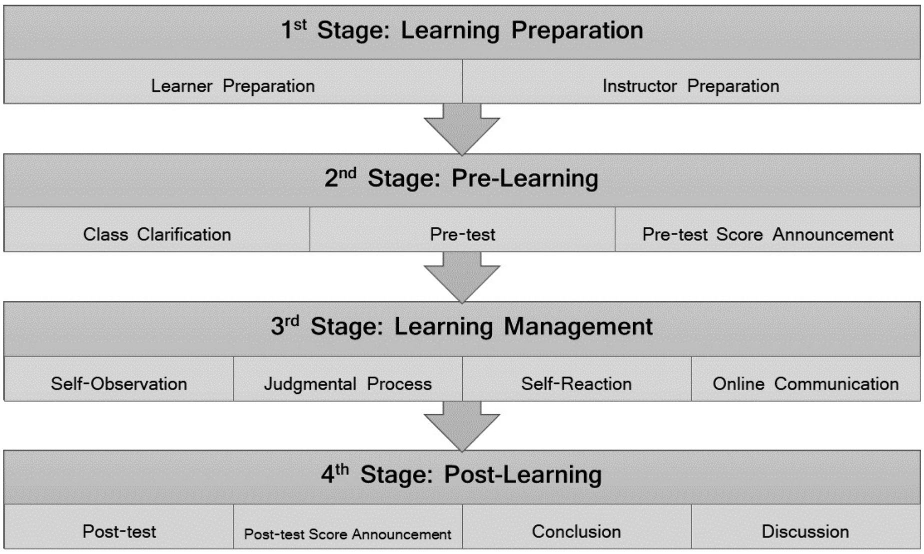

Fig. 6. Four main activities in learning process

Figure 6 shows the processes and activities that have taken place in the course 221203, Technology for Business Application at the School of Information and Communication Technology, the University of Phayao. It consists of four main stages and thirteen sub-stages:

The first stage is the learning preparation composed of learner and instructor preparation. The learner preparation process is carried out by explaining the course description, course content, learning evaluation methods, and learning resources. Instructor preparation comprises teaching planning, teaching material, and pretest and posttest.

The second stage is the pre-learning stage, which is the initial process of learning management. There are three components including class clarifying, pre-test taking and pre-test scores summarizing. In the pre-test, the exam is an online exam with randomized questions and answers. The exam has fifteen minutes as a limited time. Learners need to complete within the time limit. After the learner completed this stage, the instructor will summarize and discuss the initial score.

The third stage is the learning management phase using blended learning and self-regulated learning theory in learning management based on four steps. The self-observation step defines the learning goals for each period. The judgmental process is to consider the learning goals at the end of the class. The self-reaction step is a self-rewarding or self-punishing phase. For example, if a learner gets a post-test score less than the pre-test score, they will commit themselves to resolving it in the next exam. The last step is online communication where the instructor allows students to communicate with them either synchronously or asynchronously through online processes, such as chat, phone calls, email, etc.

After the learning management process is completed, the final step will assess the knowledge of learners who participate in class activities. This stage is the post-learning phase. The components in this process consist of four parts: taking the post-test, summarizing the post-test scores, summarizing the assessment results, and discussing the class results. The four activities in this class lead to the collection of data that will be used for analysis to find suitable learning models and to further discover the factors that support learning achievement.

3.3 Data preparation

This phase is the process of converting the collected data (raw data) into the formatted data that can be analyzed in the modeling phase. It consists of five sub-steps: select, clean, construct, integrate, and format data [39].

In this research, data preparation is a collection of the designed data as shown in Figures 5 and 6. Data gleaned from the learning process-based approach to blended learning and self-regulated learning with self-regulated learning strategies through online learning management. The attribute's collected data are listed by the categories, shown in Table 1.

Table 1. Collection of attributes classified by category

Blended Learning Strategy	Self-Regulated Learning Strategy	Academic Scores and Academic Results
• *The number of class attendances*: There are fifteen learning activities.	• *Self-Observation Phase*: It uses pre-test scores as a goal-setting tool. The goal is to have an increase in post-test scores of more than or equal to 30%.	• *Academic Scores:* ○ Quiz scores, ○ Midterm score, ○ Final score
• *The number of participants in the pre-test.*	• *Judgmental Process Phase*: It uses post-test scores as a goal comparison tool.	• *Academic Results (Criteria):* ○ 80.00–100 = A ○ 75.00–79.99 = B+ ○ 70.00–74.99 = B ○ 65.00–69.99 = C+ ○ 60.00–64.99 = C ○ 55.00–59.99 = D+ ○ 50.00–54.99 = D ○ 0.00–49.99 = F
• *The number of participants in the post-test.*	• *Self-Reaction Phase*: It is classified into two types: Positive Reward, Negative Reward	

Note: *It contains a total of eight pre-test and post-test activities.

Table 2. Examples of collected data

SID	Blended Learning Strategy			Self-Regulated Learning Strategy							Academic Scores and Academic Results			
	nAtten	nPre	nPost	Pre1	Post1	ReAct1	…	Pre8	Post8	ReAct8	Quiz Score	Midterm Score	Final Score	Academic Result
Instance 01	15	8	8	9	9	Positive	…	3	10	Positive	28.50	17.00	24.00	B
Instance 02	14	7	7	2	8	Positive	…	3	10	Positive	24.60	18.00	30.00	B
Instance 03	15	8	8	8	8	Positive	…	3	10	Positive	27.00	19.00	27.00	B
Instance 04	15	8	8	4	8	Positive		4	10	Positive	26.70	17.00	26.50	B
Instance 05	15	6	6	n/a	9	Positive		6	10	Positive	21.30	20.00	29.88	B
Instance 06	13	8	8	3	9	Positive		1	10	Positive	28.50	22.00	36.00	A
Instance 07	15	7	8	n/a	6	Positive		4	10	Positive	22.20	18.00	25.00	C+
Instance 08	15	7	8	n/a	9	Positive		4	10	Positive	28.20	19.00	26.00	B
Instance 09	15	8	8	3	9	Positive		4	10	Positive	28.50	19.00	28.50	A
Instance 10	15	7	8	n/a	9	Positive		2	10	Positive	26.70	17.00	34.50	B+
Instance 11	15	7	8	n/a	7	Positive		4	9	Positive	26.70	20.00	39.50	A
Instance 12	14	7	7	6	8	Positive		3	10	Positive	27.90	19.00	29.50	B+
Instance 13	14	8	6	9	9	Negative		4	10	Positive	28.50	20.00	23.00	B
…	…	…	…	…	…	…		…	…	…	…	…	…	
Instance 26	14	6	6	n/a	2	Negative		3	9	Positive	12.00	19.00	29.00	C

Table 1 presents a collection of attributes classified by category, which are classified into three categories: blended learning strategy, self-regulated learning strategy, and academic scores and results.

Attributes in the blended learning strategy contain three attributes that imply the cooperation of the learners in the activities. It consists of several class attendances, the number of participants in the pre-test, and the number of participants in the post-test. The learning activities in this course consist of 15 lectures and activities on an online management system. The pre-test and post-test activities are the controlled activities aligned with the four main activities in the learning process as shown in Figure 6. It has organized eight activities to make the learning process with the designed research process.

Attributes in the self-regulated learning strategy are three attributes that govern learners in all activities. Self-observation phase uses pre-test scores as a goal-setting tool. The goal is to have an increase in post-test scores of more than or equal to 30%. The judgmental process phase uses post-test scores as a goal comparison tool. Self-reaction phase is classified into two types including positive and negative rewards.

Attributes in the academic scores and results are four attributes that determine the learner's academic achievement. It consists of quiz scores, midterm scores, final scores, and academic results. Examples of collected data are shown in Table 2.

3.4 Modeling

The modeling phase is the stage of data analysis to find appropriate models from the prepared data. There are four processes including selecting modeling techniques, generating test designs, building models, and assessing the model.

This modeling aims to study the insight learning models from educational data mining techniques for online learning management. The comparison modeling process was designed by pre-midterm and post-midterm activities. It comes up with the developed models referred to the qualifications from Table 1 for model comparison. The process of creating and comparing models is shown in Figure 7.

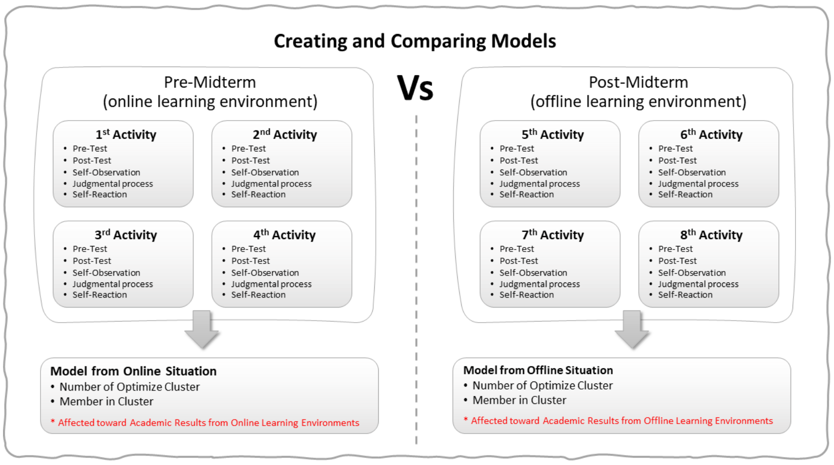

Fig. 7. The process of creating and comparing models

Figure 7 shows the research elements to identify significant factors for the deployment of the self-regulated learning theory in managing online learning. In this modeling, the researcher wanted to study the correlation of the appropriate cluster for academic achievement. The comparisons used for this study were determined with the same sample target in different situations and models. In this research, the researcher was highly anticipated discovering the differences, similarities, and correlations of patterns that change the situation.

The tool used to cluster learners in this process is data mining. The researcher applied the k-means as an appropriate clustering technique. The k-means clustering method is one of the most widely accepted grouping methods based on the concept of unsupervised learning in machine learning. The target of k-means is to identify the behavior of similar data and cluster them together. At the same time, it separates the distant data to create another cluster. The similarity calculation concept can be determined by a Euclidean distance or a simple line between two points. The shorter the Euclidean distance, the score indicates the appropriateness of the clustering of the points.

The k-means algorithm consists of 4 steps. Step 1 is to determine the required number of clusters. Step 2 is to select a random point for each centroid. Notice that the number of centroids is equal to the number of selected clusters. Step 3 is to calculate the Euclidean distance, calculated between each point and centroid. The calculated point is assigned to the nearest centroid of the cluster by Euclidean distance. Each data point can reside in only one cluster. In this step, the result in the centroid will be changed. Step 4 is to consider the centroid. If there is any change, repeat step 3; however, in the case that there is no change, stop working on any processes.

The major challenge of the k-means clustering is to find the optimal k value. The reason is that a good selection of k values is consistent with the characteristics and behavior of cluster members. Therefore, the following sections will be presented for consideration to select appropriate values of k.

3.5 Evaluation

In this step, the results of the data analysis using the data mining technique are presented. However, before the results are put into operation, it is important to measure the effectiveness of the results to match the objectives set in the first step.

The results of both evaluation and assessment were aimed at selecting appropriate clusters to determine the effectiveness of the selected clusters. The theory used to determine the appropriate number of clusters applies a method, known as k-Determination [13]. It uses the elbow principle considered by instantaneous change. The elbow method is a very popular iterative statistical technique for determining the optimal cluster count by running the k-means algorithm on a range of cluster values. The elbow method calculates the sum of the square distances from each point to the centroid given for each iteration of the k-means. Each iteration runs through a different number of clusters. The result is a line chart showing the sum of the square distances in each cluster. The result is the acquisition of several segments through structured decisions. The researcher used this theory to compare results in different situations.

3.6 Deployment

The workflow of CRISP-DM does not stop with the results obtained from the analysis of the data using data mining techniques. Although the results show some useful knowledge. But the knowledge gained must be applied to the organization or company. In the deployment phase, it must take the results of the evaluation and define a strategy for the deployment. If a general procedure is specified to create an associated model, it will be recorded here for later deployment. Therefore, it is important to consider approaches and methods of application in understanding the research problems. Therefore, it consists of four subsections of deployment including plan deployment, plans monitoring and maintenance, producing a final report, and reviewing the project.

In this research, the researcher used the appropriate cluster analysis to design the learning style that is consistent with the learners' behavior in unusual situations. Additionally, the researcher strongly believes that this research can uncover new learner skills and new learning styles judicious for 21st-century learning. The researcher has a plan to deploy as follows:

Deployment Plan: Outline the deployment strategy including the necessary steps and actions.

Monitoring and Maintenance Plan: Summarize the monitoring and maintenance strategy including the necessary steps and performances.

Produce Final Report: Report the data mining involvement. It includes all previous deliverables, summaries, and the organization of results.

Review Project: Summarize the important experiences gained during the research.

4 Research results

4.1 Population and sample

The population is the students who enrolled in the course 221203, Technology for Business Application in the 2nd semester of the academic year 2020 at the School of Information and Communication Technology, the University of Phayao, Thailand.

The determination of sample size is the selection of purposive sampling. It is a sample selection based on the decision of the researcher. The nature of the group selected the purposes of the research. The sample group was 26 students who enrolled in course 221203 Technology for Business Application in the 2nd semester of the academic year 2020 at the School of Information and Communication Technology, the University of Phayao, Thailand. In conclusion, the data based on the sample is shown in Tables 3 and 4.

Table 3. Data collection analysis

Gender	Quiz Score	Midterm Score	Final Score	Academic Results								Total
				A	B+	B	C+	C	D+	D	F	
Male	24.03	18.21	29.83	5	3	7	2	1	0	0	1	19
Female	27.51	19.29	29.36	2	2	3	0	0	0	0	0	7
Average	24.97	18.50	29.71	7	5	10	2	1	0	0	1	26

Table 3 shows the analysis of the collected data consisting of quiz scores, midterm scores, final scores, and academic results. The collected data contained more male than female students (19:7). However, overall scores of females were slightly higher than males. In the academic results dimension, females had higher grades than males compared to the amount of gathered data.

Table 4. Average activity scores

Gender	Activity Scores								Average	
Average Pre-Midterm Activity Scores										
	Pre-1	Post-1	Pre-2	Post-2	Pre-3	Post-3	Pre-4	Post-4	Pre	Post
Male	2.26	7.53	3.68	7.79	4.47	7.58	2.74	7.47	3.29	7.59
Female	4.86	9.00	5.86	9.85	6.00	9.43	4.43	9.71	5.29	9.50
Average	2.96	7.92	4.27	8.35	4.89	8.08	3.19	8.08	3.83	8.11
Average Post-Midterm Activity Scores										
	Pre-5	Post-5	Pre-6	Post-6	Pre-7	Post-7	Pre-8	Post-8	Pre	Post
Male	5.26	6.63	3.63	8.68	3.58	8.79	3.63	8.37	4.03	8.12
Female	5.86	9.71	5.43	10.00	4.43	9.71	3.43	9.71	4.79	9.78
Average	5.42	7.46	4.12	9.04	3.81	9.04	3.58	8.73	4.23	8.57

Note: Pre = Pre-test, Post = Post-test, Maximum score = 10 scores.

Table 4 shows the average of the activities in each section. It is classified into two parts as shown in the pre-midterm activity score and the post-midterm activity score. According to Table 4, it was found that female learners had higher learning advantages than males. It can be determined by the average across all activities.

4.2 Model's analysis and evaluation

The objective of the model's analysis and evaluation was to study insight learning models from educational data mining techniques for online learning management. The scope of the modeling process is presented in Figure 7. It consists of two parts: pre-midterm and post-midterm activities. The four-stage elements in all activities are the same as shown in Figure 6. There are learning preparation, pre-Learning, learning management, and post-Learning stage.

In conclusion, the attributes used in the analysis of the model are pre-test scores and post-test scores for constructing the appropriate clusters as shown in Figures 8 to 9 and detailed in Tables 5 to 8. Finally, members of each cluster will be considered for comparative academic results within the cluster as shown in Tables 6 and 8.

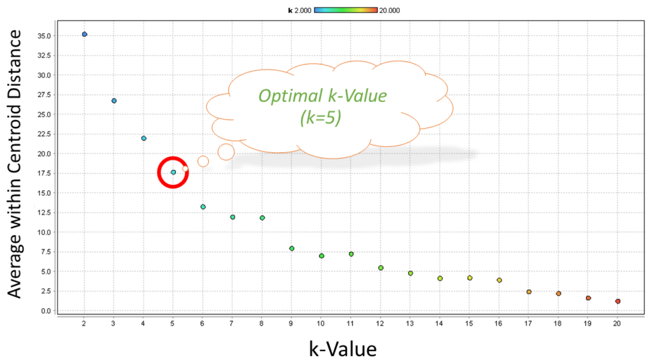

Fig. 8. Optimal k-Value from pre-midterm activities

Figure 8 shows the selection of the appropriate k-Value based on k-Determination from pre-midterm activities. It was discovered that the k-Value that should be used for the next part of the research was k equal to 5. In addition, the researcher presented the centroids of each attribute as shown in Table 5, and Table 6 shows members in each cluster.

Table 5. Centroids from pre-midterm activities analysis

Attributes	Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
1st Pre-Test Score	0.00	0.00	2.00	5.62	0.29
1st Post-Test Score	1.00	7.67	8.00	8.69	8.57
2nd Pre-Test Score	3.00	3.67	3.00	4.38	4.86
2nd Post-Test Score	0.00	7.67	7.00	9.31	9.43
3rd Pre-Test Score	2.00	8.33	0.00	4.92	4.86
3rd Post-Test Score	0.00	6.33	0.00	10.00	8.71
4th Pre-Test Score	0.50	1.00	2.00	3.54	4.43
4th Post-Test Score	1.50	2.00	10.00	9.77	9.14

Table 5 shows the centroids from the pre-midterm activities analysis. It consists of five clusters, with each cluster member showing a correlation to the learner's achievement. It is presented in Table 6.

Table 6. Correlation between cluster members and academic results

Cluster	Academic Results								Members
	A	B+	B	C+	C	D+	D	F	
Cluster_0	0	0	0	0	1	0	0	1	2
Cluster_1	0	0	1	2	0	0	0	0	3
Cluster_2	0	0	1	0	0	0	0	0	1
Cluster_3*	5	2	6	0	0	0	0	0	13
Cluster_4	2	3	2	0	0	0	0	0	7
Total	7	5	10	2	1	0	0	1	26

Table 6 shows the correlation between cluster members and academic results. It was discovered that most of the members were in cluster_3. Most of the members had high academic results within this cluster. It can be concluded that online learning activities (pre-midterm activities) can be categorized into 5 clusters. Thus, the cluster with the most members and the high learning outcome was cluster_3.

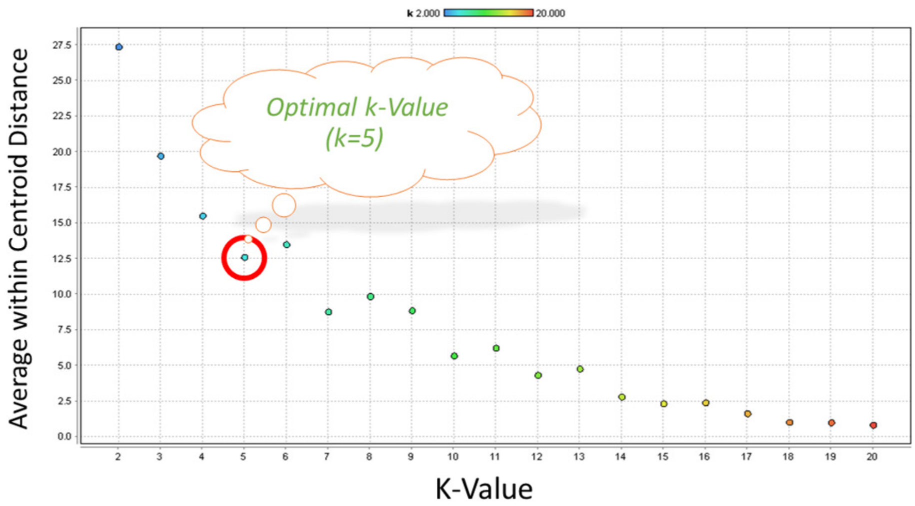

Fig. 9. Optimal k-value from post-midterm activities

Figure 9 shows the selection of the appropriate k-Value based on k-Determination from post-midterm activities. It was discovered that the k-Value that should be used for the next part of the research where k was equal to 5. In addition, the researcher presented the centroids of each attribute as shown in Table 7. Moreover, Table 8 shows members in each cluster.

Table 7. Centroids from pre-midterm activities analysis

Attributes	Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
5th Pre-Test Score	5.60	8.00	7.00	2.50	4.50
5th Post-Test Score	9.50	0.00	0.00	0.00	2.00
6th Pre-Test Score	4.45	0.00	0.00	6.00	3.00
6th Post-Test Score	9.95	0.00	0.00	8.50	9.50
7th Pre-Test Score	3.90	0.00	3.00	4.50	4.50
7th Post-Test Score	9.55	0.00	10.00	7.50	9.50
8th Pre-Test Score	3.85	0.00	3.00	3.50	3.00
8th Post-Test Score	9.80	0.00	10.00	9.50	1.00

Table 7 shows the centroids from the pre-midterm activities analysis. It consists of five clusters. Each cluster member shows a correlation to the learner's achievement. It is presented in Table 8.

Table 8. Correlation between cluster members and academic results

Cluster	Academic Results								Members
	A	B+	B	C+	C	D+	D	F	
Cluster_0*	6	3	10	1	0	0	0	0	20*
Cluster_1	0	0	0	0	0	0	0	1	1
Cluster_2	0	1	0	0	0	0	0	0	1
Cluster_3	1	0	0	1	0	0	0	0	2
Cluster_4	0	1	0	1	0	0	0	0	2
Total	7	5	10	2	1	0	0	1	26

Table 8 shows the correlation between cluster members and academic results. It was discovered that most of the members were in cluster_0. Most of the members had high academic results within this cluster. It can be concluded that online learning activities (pre-midterm activities) can be categorized into 5 clusters. Thus, the cluster with the most members and the high learning outcome was cluster_0. The next section discusses the research findings based on the research objectives.

5 Research discussion

The discussion in this section is defined by two main objectives: (1) to study the insight learning model from educational data mining techniques for online learning management, and (2) to study the learning achievement factors of learners applying blended learning and self-regulated learning.

Note that the learning model refers to the provision of learning to achieve academic achievement.

5.1 Insight learning model

According to the modeling process, shown in Figure 7, there are two main processes for classifying and comparing models. It consists of pre-midterm and post-midterm activities.

The reasonable clustering process for k-Value from pre-midterm activities is illustrated in Figure 8. A detailed description of the centroid values for each cluster is given in Table 5. In addition, the members and the relationship of the learners' academic achievement from pre-midterm activities are shown in Table 6. On the other hand, the judicious clustering process for k-Value from post-midterm activities is demonstrated in Figure 9. An explanation of the centroid values for each cluster is shown in Table 7. Additionally, the members and the relationship of the learners' academic achievement from post-midterm activities are shown in Table 8.

Based on the research results, two points can be drawn up as follows. (1) The optimal number of clusters for both online and offline learning was not different. It can be compared in Figures 8 and 9. (2) The number of cluster members will change as the learning management model or learning environment changes. In this dimension, the researcher discovered that learners were clustered according to their learning environment and activity. The evidence is clearly shown in Tables 6 and 8.

However, it may be questioned how pre-midterm and post-midterm cluster members have dispersed or changed. Therefore, the researcher summarizes the members and clusters altered from pre-midterm and post-midterm activities in Table 9.

Table 9. Members and clusters altered from pre-midterm and post-midterm activities

Instance	Gender	Academic Results	Pre-Midterm Cluster	Post-Midterm Cluster
Instance 01	Female	B	Cluster_3	Cluster_0
Instance 02	Male	B	Cluster_2	Cluster_0
Instance 03	Male	B	Cluster_3	Cluster_0
Instance 04	Male	B	Cluster_3	Cluster_0
Instance 05	Male	B	Cluster_1	Cluster_0
Instance 06	Male	A	Cluster_3	Cluster_0
Instance 07	Male	C+	Cluster_1	Cluster_0
Instance 08	Male	B	Cluster_4	Cluster_0
Instance 09	Male	A	Cluster_3	Cluster_0
Instance 10	Female	B+	Cluster_4	Cluster_0
Instance 11	Male	A	Cluster_4	Cluster_3
Instance 12	Male	B+	Cluster_3	Cluster_2
Instance 13	Female	B	Cluster_3	Cluster_0
Instance 14	Female	A	Cluster_4	Cluster_0
Instance 15	Male	A	Cluster_3	Cluster_0
Instance 16	Male	F	Cluster_0	Cluster_1

(Continued)

Table 9. Members and clusters altered from pre-midterm and post-midterm activities *(Continued)*

Instance	Gender	Academic Results	Pre-Midterm Cluster	Post-Midterm Cluster
Instance 17	Male	B+	Cluster_4	Cluster_4
Instance 18	Female	B+	Cluster_3	Cluster_0
Instance 19	Female	A	Cluster_3	Cluster_0
Instance 20	Male	B	Cluster_3	Cluster_0
Instance 21	Male	C+	Cluster_1	Cluster_4
Instance 22	Male	B	Cluster_3	Cluster_0
Instance 23	Male	A	Cluster_3	Cluster_0
Instance 24	Male	B+	Cluster_4	Cluster_0
Instance 25	Female	B	Cluster_4	Cluster_0
Instance 26	Male	C	Cluster_0	Cluster_3

Table 9 shows the changing members and clusters of pre-midterm and post-midterm activities. When considering Cluster_3 from pre-midterm activity analysis and Cluster_0 from post-midterm activity analysis, it seems that these two clusters are extremely similar by comparison to the centroid. However, some members are not in this relationship.

Based on these findings, the researcher studied the latent factors that influence the learning model as discussed in the next section.

5.2 Learning achievement factors

The attributes mentioned in the research are presented in Table 1. It consists of three main categories including blended learning strategy attributes, self-regulated learning strategy attributes, and attributes of academic scores and academic results.

Based on the results of the research, the researcher found that activity participation was related to classroom activity. The collected data in Table 2 and the collected data are presented on the website: https://bit.ly/3u0u5T5.

The significant finding is that learners place more emphasis on post-test activities than pre-test activities. The important evidence is the sixth, seventh, and eighth activities. It found that the learners had the highest scores after the activity (post-test) and had very low scores before the activity (pre-test). In addition, some learners do not agree to cooperate in pre-test and post-test activities even if they attend as normal. Moreover, the pre-midterm scores (online learning) are lower than the post-midterm scores (offline learning). It can therefore be concluded that learners still need close supervision even if they are in the higher education system.

In summary, the factors of blended learning and self-regulated learning benefit learners' support of the learning model. It is suggested by the researcher that the students still need control and supervision at a close level. Instructors should have activities that focus learners on the learning goals of each class. Finally, the instructors should also use the findings of this research as a guideline for designing an appropriate learning model.

6 Conclusion

The COVID-19 pandemic has forced a sudden change in the educational process in Thailand. Universities were restricted to providing an online learning model only. The consequences have a severe impact on the learners in raising boresome in the online learning style. Therefore, this research aims to study the paradigm shift of the education system focusing on the deployment of artificial intelligence technology to support the learning model during the COVID-19 pandemic situation. The purpose of the research is (1) to study an appropriate self-regulated learning model with data mining techniques for designing appropriate online learning management, and (2) to study the learning achievement factors of learners by applying blended learning and self-regulated learning techniques. The samples were 26 students who enrolled in the course 221203, Technology for Business Application in the 2nd semester of the academic year 2020 at the University of Phayao. The research tool was statistical analysis and a machine learning tool. It consists of analyzing pre-test scores, post-test scores, midterm scores, final scores, academic achievement, clustering analysis, and clustering performance.

As a result, the researcher achieved the stated research objectives. This study summarized the results and discussed the following key points. The first point indicated that there was no difference in the analysis of the student clustering of the two time periods comprising before midterm (pre-midterm) and after midterm (post-midterm). Although the researcher encourages teaching with different teaching methods and techniques, the findings from the questionnaire to assess learning management after the course, it found that the learners had been bored with online teaching for the past two years. In addition, learners need educational interaction and practice in the university's laboratories.

The second point demonstrated that the researcher wanted to study the factors that drive and motivate the learners. The researcher found that the self-regulated learning theory obtained gaining popularity and recognition. However, it was not appropriate for learners to control and set goals for their learning entirely without the involvement of an instructor. Moreover, the general online learning process was a major obstacle to setting learning goals as mentioned in the research discussion section.

However, the analysis result is the selection of five clusters suitable for both environments in online and offline learning in Figures 8 and 9. A detailed description of the centroid values for each cluster is given in Tables 5 and 7. In addition, the members and the relationship of the learners' academic achievement from pre-midterm and post-midterm activities are shown in Tables 6 and 8. Finally, it can be concluded that there was a little difference in organizing the learning online and offline. Therefore, in future research, the researcher needs to perform research in the scope of identifying the suitability and necessity of converting the face-to-face learning model to a fully online learning model. In addition, other teaching techniques and methods should be discovered to motivate students.

7 Acknowledgment

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao (Grant No. FF65-UoE006). In addition, this research was supported by many advisors, academics, researchers, students, and academic staff at the School of Information and Communication Technology, the University of Phayao. The author would like to thank all of them for their support and collaboration in making this research.

8 Conflict of interest

The author declares no conflict of interest.

9 References

[1] M. Flavin, "Disruptive technologies in higher education," *Res. Learn. Technol.*, vol. 20, 2012. https://doi.org/10.3402/rlt.v20i0.19184

[2] S. Gallagher and G. Garrett, "Disruptive education: Technology-enabled universities," *U. S. Stud. Cent. Univ. Syd.*, 2013.

[3] S. A. Stage and D. R. Quiroz, "A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings," *Sch. Psychol. Rev.*, vol. 26, no. 3, pp. 333–368, Sep. 1997. https://doi.org/10.1080/02796015.1997.12085871

[4] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, p. e1355, 2020. https://doi.org/10.1002/widm.1355

[5] W. Nuankaew and P. Nuankaew, "The study of the factors and development of educational model: The relationship between the learner context and the curriculum context in higher education," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 21, pp. 205–226, Nov. 2019. https://doi.org/10.3991/ijet.v14i21.11034

[6] E. Araka, E. Maina, R. Gitonga, and R. Oboko, "A conceptual model for measuring and supporting self-regulated learning using educational data mining on learning management systems," in *2019 IST-Africa Week Conference (IST-Africa)*, May 2019, pp. 1–11. https://doi.org/10.23919/ISTAFRICA.2019.8764852

[7] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, Jan. 2015. https://doi.org/10.1016/j.procs.2015.12.157

[8] A. S. Goodsell, "Collaborative learning: A sourcebook for higher education," 1992.

[9] B. L. Smith and J. T. MacGregor, "What is collaborative learning," *Virtual Univ. Int. Online Learn. Perspect.*, pp. 217–232, 1992.

[10] P. Appiah-Kubi and E. Annan, "A review of a collaborative online international learning," *Int. J. Eng. Pedagogy IJEP*, vol. 10, no. 1, Art. no. 1, Jan. 2020. https://doi.org/10.3991/ijep.v10i1.11678

[11] P. Dillenbourg, *What Do You Mean by Collaborative Learning?* Oxford: Elsevier, 1999.

[12] J. van der Linden, G. Erkens, H. Schmidt, and P. Renshaw, "Collaborative learning," in *New learning*, Springer, 2000, pp. 37–54. https://doi.org/10.1007/0-306-47614-2_3

[13] P. Nuankaew and P. Temdee, "Matching of compatible different attributes for compatibility of members and groups," *Int. J. Mob. Learn. Organ.*, vol. 13, no. 1, pp. 4–29, 2019, https://doi.org/10.1504/IJMLO.2019.096469

[14] F. J. García-Peñalvo, Á. Fidalgo-Blanco, and M. L. Sein-Echaluce, "An adaptive hybrid MOOC model: Disrupting the MOOC concept in higher education," *Telemat. Inform.*, vol. 35, no. 4, pp. 1018–1030, Jul. 2018. https://doi.org/10.1016/j.tele.2017.09.012

[15] F. Mossavar-Rahmani and C. Larson-Daugherty, "Supporting the hybrid learning model: A new proposition," *MERLOT J. Online Learn. Teach.*, vol. 3, no. 1, pp. 67–78, 2007.
[16] B. Collis and J. Moonen, "Flexible learning in a digital world," *Open Learn. J. Open Distance E-Learn.*, vol. 17, no. 3, pp. 217–230, 2002. https://doi.org/10.1080/0268051022000048228
[17] K. Chaharbaghi and V. Newman, "Innovating: Towards an integrated learning model," *Manag. Decis.*, 1996. https://doi.org/10.1108/00251749610115107
[18] A. Gajda, R. A. Beghetto, and M. Karwowski, "Exploring creative learning in the classroom: A multi-method approach," *Think. Ski. Creat.*, vol. 24, pp. 250–267, 2017. https://doi.org/10.1016/j.tsc.2017.04.002
[19] E. Y. Huang, S. W. Lin, and T. K. Huang, "What type of learning style leads to online participation in the mixed-mode e-learning environment? A study of software usage instruction," *Comput. Educ.*, vol. 58, no. 1, pp. 338–349, 2012. https://doi.org/10.1016/j.compedu.2011.08.003
[20] T. Anderson, *The theory and practice of online learning*. Athabasca University Press, 2008.
[21] K. E. Rudestam and J. Schoenholtz-Read, *Handbook of online learning*. Sage Publications, 2009.
[22] Z. I. Almarzooq, M. Lopes, and A. Kochar, "Virtual learning during the COVID-19 pandemic: A disruptive technology in graduate medical education," *J. Am. Coll. Cardiol.*, vol. 75, no. 20, pp. 2635–2638, May 2020. https://doi.org/10.1016/j.jacc.2020.04.015
[23] C. M. Christensen, M. E. Raynor, and R. McDonald, "What is disruptive innovation," *Harv. Bus. Rev.*, vol. 93, no. 12, pp. 44–53, 2015.
[24] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, "Intelligence unleashed: An argument for AI in education," 2016.
[25] M. Billinghurst, "Augmented reality in education," *New Horiz. Learn.*, vol. 12, no. 5, pp. 1–5, 2002.
[26] I. Plauska and R. Damaševičius, "Educational robots for internet-of-things supported collaborative learning," in *International Conference on Information and Software Technologies*, 2014, pp. 346–358. https://doi.org/10.1007/978-3-319-11958-8_28
[27] M. Ozkana and B. Solmazb, "Mobile Addiction of Generation Z and its Effects on their Social Lifes: (An Application among University Students in the 18–23 Age Group)," *Procedia – Soc. Behav. Sci.*, vol. 205, pp. 92–98, Oct. 2015. https://doi.org/10.1016/j.sbspro.2015.09.027
[28] A. Bandura, "Social-learning theory of identificatory processes," *Handb. Social. Theory Res.*, vol. 213, p. 262, 1969.
[29] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Psychol. Rev.*, vol. 84, no. 2, p. 191, 1977. https://doi.org/10.1037/0033-295X.84.2.191
[30] K. Bussey and A. Bandura, "Social cognitive theory of gender development and differentiation," *Psychol. Rev.*, vol. 106, no. 4, p. 676, 1999. https://doi.org/10.1037/0033-295X.106.4.676
[31] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," *Front. Psychol.*, vol. 8, p. 422, 2017. https://doi.org/10.3389/fpsyg.2017.00422
[32] M. Boekaerts, M. Zeidner, and P. R. Pintrich, *Handbook of self-regulation*. Elsevier, 1999. https://doi.org/10.1016/B978-012109890-2/50030-5
[33] B. J. Zimmerman and A. R. Moylan, "Self-regulation: Metacognition and motivation intersect," 2009.
[34] B. J. Zimmerman, "From cognitive modeling to self-regulation: A social cognitive career path," *Educ. Psychol.*, vol. 48, no. 3, pp. 135–147, Jul. 2013. https://doi.org/10.1080/00461520.2013.794676

[35] R. A. Rasheed, A. Kamsin, and N. A. Abdullah, "Challenges in the online component of blended learning: A systematic review," *Comput. Educ.*, vol. 144, p. 103701, Jan. 2020. https://doi.org/10.1016/j.compedu.2019.103701

[36] C. Sahin, "Social media addiction scale-student form: The reliability and validity study," *Turk. Online J. Educ. Technol.-TOJET*, vol. 17, no. 1, pp. 169–182, 2018. https://doi.org/10.1037/t72756-000

[37] L. Scahill and M. Schwab-Stone, "Epidemiology of ADHD in school-age children," *Child Adolesc. Psychiatr. Clin. N. Am.*, vol. 9, no. 3, pp. 541–555, Jul. 2000. https://doi.org/10.1016/S1056-4993(18)30106-8

[38] P. Nuankaew, "Dropout situation of business computer students, University of Phayao," *Int. J. Emerg. Technol. Learn. IJET*, vol. 14, no. 19, Art. no. 19, Oct. 2019. https://doi.org/10.3991/ijet.v14i19.11177

[39] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.

10 Author

Pratya Nuankaew is currently an instructor at the School of Information and Communication Technology, University of Phayao, Phayao, 56000, Thailand. (Email: pratya.nu@up.ac.th) He is the corresponding author of this research. His research interests are applied informatics technologies, behavioral sciences analysis with technologies, computer-supported collaborative learning, data science in education, educational data mining, learning analytics, and learning styles, learning strategies for lifelong learning, self-regulated learning, social network analysis, and ubiquitous computing.

Article submitted 2021-04-28. Resubmitted 2022-05-30. Final acceptance 2022-05-31. Final version published as submitted by the authors.