

# Application for Identifying Students Achievement Prediction Model in Tertiary Education

## Learning Strategies for Lifelong Learning

<https://doi.org/10.3991/ijim.v15i22.24069>

Pratya Nuankaew<sup>1(✉)</sup>, Patchara Nasa-ngium<sup>2</sup>, Wongpanya Sararat Nuankaew<sup>2</sup>

<sup>1</sup>University of Phayao, Phayao, Thailand

<sup>2</sup>Rajabhat Maha Sarakham University, Maha Sarakham, Thailand

pratya.nu@up.ac.th

**Abstract**—The purpose of the research is to identify the risk of dropping out in tertiary students with an application. The components of the research goal aim (1) to develop the students' achievement prediction model and (2) to construct a prototype application for the predictions of the tertiary students dropping out. The research tools consisted of three parts, (1) tool for developing predictive prototypes uses a tool called the CRISP-DM process with Decision Tree Classification, Feature Selection methods, Confusion Matrix performance, Cross-Validation methods, Accuracy, Precision and Recall measurements, (2) tool for application development used the SDLC with V-method, and (3) tool to assess application satisfaction used questionnaires and statistical analysis. Data sample were collected from 401 students enrolled in the Business Computer Program at the School of Information and Communication Technology, University of Phayao during the academic year 2012–2016. The results showed that the prediction model had a very high percentage of accuracy (82.29%). The prototype test results with the data gathered had a very high score level (84.04%; correct 337 out of 401 training examples). An overview of the underlying application with the utmost integrity by the researchers planned to put the application to the test in the first semester of the academic year 2021 at the School of Information Technology and Communication, University of Phayao. For future research, the researchers plan to create a mobile application for mentors in the University of Phayao to monitor learner on both Android and iOS systems.

**Keywords**—learning analytics, dropping out, educational data mining, eruptive technology, disruptive technology

## 1 Introduction

Nowadays, educational technology is emerging at a rapidly increasing acceleration. Technology has made an impact on education in many dimensions, including its positive and negative aspects. Issues that technology support education are supporting students from non-traditional backgrounds by adopting a universal design approach [1],

supporting distance learners [2], increased internationalization, delivery of education content on a large scale whilst maintaining support for the individual learner [3]–[5], automation of administrative and academic processes [6], support for academics to create and deliver content suited for all learners [7], and so on. In opposite, technology also deteriorates the quality of education, for example, students have lower concentration and interest in learning [8], the rising cost of education, attempts to use technology for exam cheating, mobile addicted [9], [10], learners get disconnected from the real world, and many others.

Both positive and negative impacts affect the normal educational process. In case, the Business Computer Program at the School of Information and Communication Technology, University of Phayao has long been supporting government education by providing services for the learning management of the curriculum from 2001 to the present (2021). The serious problem with the curriculum is that the number of students has decreased dramatically, approximately 90 percent from the academic year 2003, with 251 students being 23 students in the academic year 2020 as shown in Figure 3. Moreover, the rate of non-graduation or drop out during the study is a constant rate with a high level of approximately 40% as presented in Figure 3, Table 1 and Table 2.

As for the importance of the above, the researchers aim to develop research with two important goals: (1) to develop the students' achievement prediction model and (2) to construct a prototype application for the predictions of the tertiary students dropping out. The research tools consisted of three parts, (1) tool for developing predictive prototypes uses a tool called the CRISP-DM process with Decision Tree Classification, Feature Selection methods, Confusion Matrix performance, Cross-Validation methods, Accuracy, Precision and Recall measurements, (2) tool for application development used the SDLC with V-method including requirement analysis, system design, architecture design, module design, coding, unit testing, integration testing, system testing, and acceptance testing, and (3) tool to assess application satisfaction used questionnaires and statistical analysis. Data analysis was calculated from 401 students enrolled in the Business Computer Program at the School of Information and Communication Technology, University of Phayao during the academic year 2012–2016.

This research is based on assumptions and beliefs that developing effective tools can effectively solve problems affecting the curriculum. Therefore, the researchers are committed to conducting in-depth research to obtain quality research by presenting the research report according to the following structure: the 1<sup>st</sup> section presents the background and importance of the research problem. The 2<sup>nd</sup> section clarifies the research materials and research methodologies. The 3<sup>rd</sup> section is the research results, the 4<sup>th</sup> section is the research discussions, the 5<sup>th</sup> section is the research conclusion, and the last section, recommendations for future research.

## **2 Materials and methodologies**

There are three phases of research materials and methodologies: (1) data collection, (2) modeling, and (3) application construction. All details are followed the research framework as shown in Figure 1.

## Research Framework

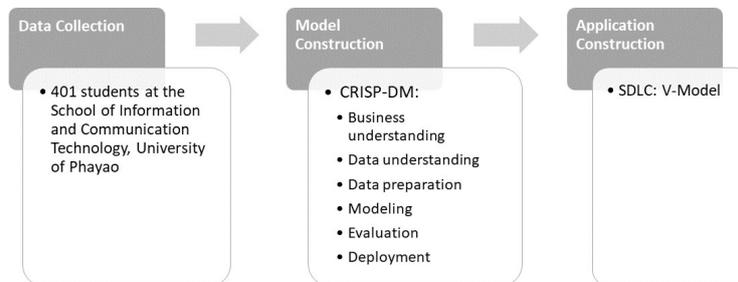


Fig. 1. Research framework

### 2.1 Data collection

The data collection was collected from 401 students at the School of Information and Communication Technology, University of Phayao. The data collected are the students' academic performance of the Business Computer program during the academic year 2012–2016. The variables of the collected data consisted of course details, academic results in each course, and academic achievement status in the Business Computer program during the academic year 2012–2016. The researchers summarize the collected data as shown in Table 1 and Table 2.

### 2.2 Modeling

The model of CRISP-DM (Cross-Industry Standard Process for Data Mining) is a popular model in the field of data mining analysis. In this research, it was used to develop the students' achievement prediction model. It has six productive work processes: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. An overview of the CRISP-DM model process is shown in Figure 2.

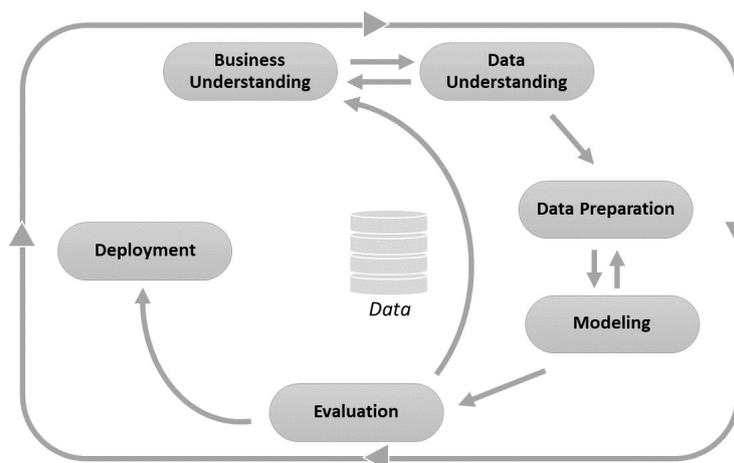


Fig. 2. CRISP-DM model

**Business understanding.** The first step is Business Understanding. The objective is to present the context of the problem with relevant goals and information so that researchers can connect the data in a business model or research question. It is composed of four sub-stages: determine business objectives, assess situation, determine data mining goals, and produce project plan [11]. Whereas the product of this process is that the research team understands the context of the research project, which aims to create a model that is significant to predict the academic achievement of students at the tertiary level.

**Data understanding.** The second step is Data Understanding. The objective is to know exactly what can be expected and what can be achieved from the data. It examines the quality of the data in several aspects such as data integrity, data fragmentation, data density, or dataset compliance values for governance and monitoring. It is composed of four sub-stages: collect initial data, describe data, explore data, and verify data quality [11].

For this step, the understanding and awareness of the research. The researcher summarizes the problems that arise in the Business Computer program at the School of Information and Communication Technology, University of Phayao.

The researchers found that the number of students enrolled in Business Computer program tended to decline significantly in relation to the declining population in Thailand. On the other hand, the number of dropout students has increased as shown in Figure 3.

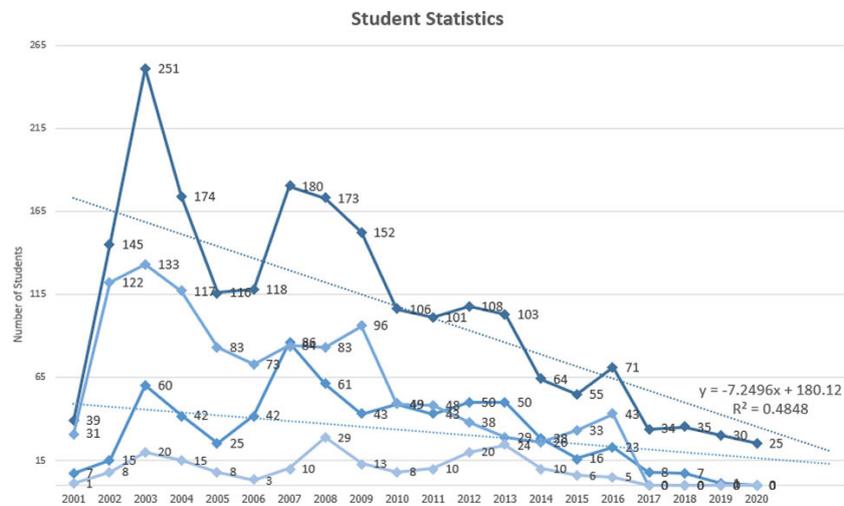


Fig. 3. Data of business computer students for twenty years (2001–2020)

Figure 3 shows the number of students who enrolled in the Business Computer program, the number of students who graduated, and the number of students who dropped out. The data collected shows information during the academic year 2001–2020. It reflects the problem of lower enrollment and the high dropout problem. However, the researchers had compiled and summarized the data, as presented in Table 1 and Table 2.

**Data preparation.** The third step is Data Preparation and involves the ETLs (Extract-Transform-Load) or ELTs (Extract-Load-Transform) process. It serves parts

and format of the data into something useful can be used in the algorithms and data mining processes for modeling. It is composed of five sub-stages: select data, clean data, construct data, integrate data, and format data [11].

For this research, data preparation starts with research ethics. Research projects are authorized under Project Code 2/020/63 on April 22, 2020 by the University of Phayao. After that, the researchers received data from the Division of Educational Services, the University of Phayao. The data received is a transaction of 254,456 students in all educational programs of the School of Information and Communication Technology from the academic year 2001 to academic year 2020.

In the next step, the researchers grouped students affiliated with the Business Computer program. It was found that there were five groups of students affiliated with the Business Computer program. The first group was 397 students enrolled during the academic year 2001–2003. The second group was 588 students enrolled during the academic year 2004–2007. The third group was 532 students enrolled during the academic year 2008–2011. The fourth group was 401 students enrolled during the academic year 2012–2016. The fifth group was 124 students enrolled during the academic year 2017–2020. The grouping is based on the program version, which is updated every 3–5 years. In the process of selecting the sample, the researchers selected a purposive sampling method by picked the sample from the fourth group, as it was the latest version with the entire process completed. The reason for not selecting the fifth group is because most of the learners are studying in the program.

From Table 1 and Table 2, it is clearly shown that most learners have dropout problems in Year 1. Therefore, the researchers extracted data on the achievement results of the learners only from the 1<sup>st</sup> semester and the 2<sup>nd</sup> semester from Year 1, which consisted of 13 courses for the analysis of educational achievement prediction. The data provided has already been shown in the link: <https://bit.ly/3xVYe8s>.

**Modeling.** The fourth step is Modeling, which is the core of any data mining project. This step is responsible for the project objectives that should meet or help achieve project goals. However, the selection of the modeling method is important, as it is necessary to select the right tools for the purpose and goals. This process is therefore made up of four sub-stages: select modeling techniques, generate test design, build model, and assess model [11].

The research consisted of two parts of the three techniques. The first part is the Decision Tree modeling techniques, which are beneficial to understand and can be utilized in a variety of ways [12]. The second part is to improve the model to be more efficient. The tool is used by the Feature Selection (FS) techniques [13], [14], it has the ability to discover the characteristics that influence the model. It can be used to reduce the number of variables for better model performance [13]. The techniques used in this part consist of two techniques: Forward Selection technique, and Backward Elimination technique. The Forward Selection technique is stacked with a subprocess. Begin the calculation by selecting an empty attribute, and each new calculation cycle adds an individual attribute of the given data set to determine model performance. In contract, the Backward Elimination technique starts the calculation with the entire set of attributes, and each round removes the existing attributes one by one and calculates them all.

All three techniques were tested in detail to determine the model's performance using confusion matrix performance, cross-validation methods, accuracy, precision and recall measurements as presented in the Evaluation section.



**Deployment.** The sixth and last step is Deployment. It contains of presenting the results in a useful way. It can be understood, and when this goal is achieved, the project should achieve it. It is composed of four sub-stages: plan deployment, plan monitoring and maintenance, produce final report, and review project [11].

This research recognizes the importance of model deployment. Therefore, the researchers have applied the most efficient model to develop into an application prototype by presenting details in the topic of application construction.

### 2.3 Application construction

In developing a good application, the developers should discover an effective of Software Development Life Cycle (SDLC). The researchers found that one of the more popular SDLCs is the V-Model [15], [16]. The V-model is a type of SDLC that has a V-shaped process to carry out activities in which all activities are assessed sequentially. It is also known as “the verification and the validation model” [15]. It consists of three phases and nine steps, as detailed and presented in this section and Figure 6.

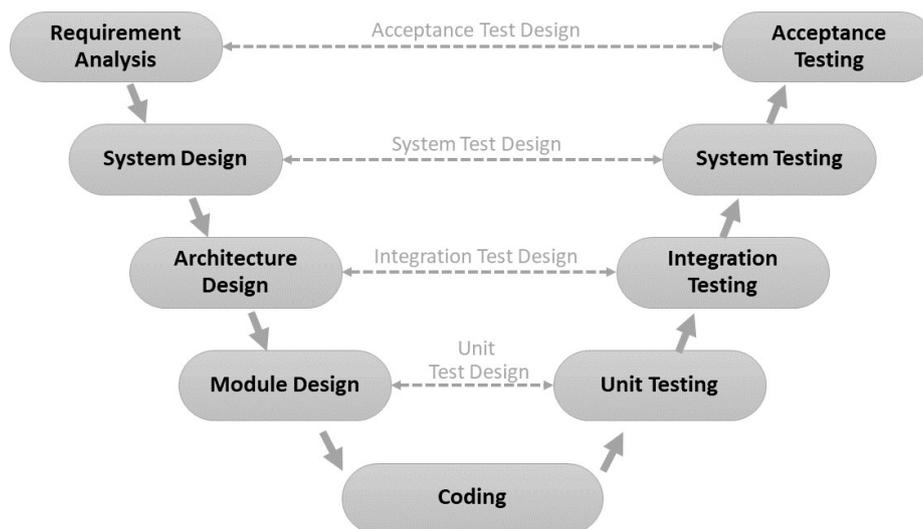


Fig. 6. SDLC: V-model [15]

Figure 6 shows the processes and steps of the V-Model, which is a type of software development life cycle model. It consists of three phases: the design phase, the coding phase, and the testing phase, as detailed below.

**Design phases.** The design phase is the preparation phase for application development. It consists of four sub-phases: requirement analysis, system design, architectural design, and module design.

For this research, the design phase is the application of a model that has been performed in the previous process to design an appropriate application. The application development process is based on the principles of application development, the details are as follows:

*Requirement Analysis:* This phase consists of detailed communication with previous studies to understand the requirement and expectations of the research problem. This step can be called “Requirement Gathering”.

*System Design:* This phase consists of the system design and hardware design setup. It considers a complete communication model for product development in which the researcher uses responsive technology for the convenience of the user.

*Architectural design:* System design is a module with different functional classification. The transfer of data and communication between the internal module and the outside world is clearly understood. In designing the program, the researcher provides the user with the pleasure of providing data and consent to the use of the application, in which the data the researcher collects in the application is based on and ethical research.

*Module Design:* At this stage, the system is divided into smaller modules, a detailed module design, known as Low-Level Design (LLD). For this research, the researcher designed an intuitive user interface with the simplest processes for the best benefit of the user.

**Coding phase.** The coding phase is a critical step, with the researchers selecting the appropriate tools for the technology and prototypes from previous analyzes. The database program used in this research was MariaDB. The computer language coding for application development is PHP, while the main framework is Laravel Framework.

**Testing phases.** The testing phase is aimed at tracking each activity that takes place in the application development. It consists of four steps: unit testing, integration testing, system testing, and user acceptance testing.

*Unit testing:* Unit testing plans are developed during the module design phase. These unit testing schemes are conducted to eliminate any code or unit level bugs.

*Integration testing:* After completion of the unit test, the integration testing is performed. To test the integration, the modules are put together and the system is tested. Integration testing will be performed in the architectural design phase. This test verifies the communication of the modules with each other.

*System testing:* Testing of a complete application testing system with functionality, interdependence, and communication. It is a test of the functional and non-functional requirements of the developed application.

*User acceptance test (UAT):* UAT is performed in a user environment similar to a production environment. UAT verifies that the delivered system meets user requirements, and that the system is world ready.

### 3 Research results

This section is the reporting of research results, divided into three parts: data collected and data sampling, academic achievement prediction model, and prototype of the prediction application model.

#### 3.1 Data collected and data sampling

The data collected is a transaction of 254,456 students in all educational programs of the School of Information and Communication Technology from the academic year

2001 to academic year 2020. In this research, the sampling scope of 401 students enrolled in the Business Computer Program at the School of Information and Communication Technology, University of Phayao during the academic year 2012–2016, as discussed in data preparation.

**Table 1.** Data collected

Academic Year	Admission	Dropped Out	Graduation	
			Scheduled	Delayed
2012	108 (26.93%)	50 (12.47%)	38 (9.48%)	20 (4.99%)
2013	103 (25.69%)	50 (12.47%)	29 (7.23%)	24 (5.99%)
2014	64 (15.96%)	28 (6.98%)	26 (6.48%)	10 (2.49%)
2015	55 (13.72%)	16 (3.99%)	33 (8.23%)	6 (1.50%)
2016	71 (17.71%)	23 (5.74%)	43 (10.72%)	5 (1.25%)
<i>Total:</i>	<i>401 (100%)</i>	<i>167 (41.65%)</i>	<i>169 (42.14%)</i>	<i>65 (16.21%)</i>

Table 1 shows the student data that was collected. It was classified annually during the academic year 2012–2016. The data collected clearly reflect the problem, with 167 high dropout students (41.65%). In addition, there are 65 students (16.21%) who fail to complete the required study period. Moreover, the huge number of students who were dropped out appears in Year 1 as shown in Table 2.

**Table 2.** Data dropped out

Academic Year	Dropped Out	Dropped Out Year						
		Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
2012	50 (29.94%)	31 (18.56%)	13 (7.78%)	2 (1.20%)	2 (1.20%)	0 (0.00%)	1 (0.60%)	1 (0.60%)
2013	50 (29.94%)	33 (19.76%)	12 (7.19%)	5 (2.99%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
2014	28 (16.77%)	19 (11.38%)	5 (2.99%)	2 (1.20%)	1 (0.60%)	0 (0.00%)	1 (0.60%)	0 (0.00%)
2015	16 (9.58%)	9 (5.39%)	5 (2.99%)	2 (1.20%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
2016	23 (13.77%)	6 (3.59%)	15 (8.98%)	1 (0.60%)	1 (0.60%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
<i>Total:</i>	<i>167 (100%)</i>	<i>98 (58.68%)</i>	<i>50 (29.94%)</i>	<i>12 (7.19%)</i>	<i>4 (2.40%)</i>	<i>0 (0.00%)</i>	<i>2 (1.20%)</i>	<i>1 (0.60%)</i>

Table 2 shows the dropped out data of students in the Business Computer program during the academic year 2012–2016. It found that 98 of the students (58.68%) dropped out during the 1<sup>st</sup> year, and 2<sup>nd</sup> year, there are 50 students (29.94%) who dropped out. From their findings, the researchers decided to select specific variables or courses in Year 1 as attributes to the analysis model. There are 13 courses as shown in Table 3.

**Table 3.** The attributes for model analysis

Code	Course Name	Credit	Type
001103	Thai Language Skills	3 (3-0-6)	GE
001111	Fundamental English	3 (3-0-6)	GE
001112	Developmental English	3 (3-0-6)	GE
003134	Civilization and Indigenous Wisdom	3 (3-0-6)	GE
004152	Body Conditioning	1 (0-2-1)	GE
005171	Life and Health	3 (3-0-6)	GE
005173	Life Skills	2 (1-2-3)	GE
122130	Principles of Management	3 (3-0-6)	Major
126100	Introduction to Economics	3 (3-0-6)	Major
128221	Principles of Marketing	3 (2-2-5)	Major
221100	Business Mathematics	3 (2-2-5)	Major
221110	Fundamental Information Technology	3 (2-2-5)	Major
221120	Introduction to Programming	3 (2-2-5)	Major

Note: GE = general education.

Table 3 shows the attributes (variables) to be used in the analysis to develop the achievement prediction model. It consists of thirteen courses in two types of courses: general education courses and major courses. These attributes were analyzed by the resulting model, which is presented in the next section.

### 3.2 Academic achievement prediction model

The models obtained from this research were presented from three techniques: Decision Tree classification technique, Forward Selection technique, and Backward Elimination technique.

**Results of the decision tree classification model.** The analysis results of the Decision Tree Classification model were presented in two dimensions. The 1<sup>st</sup> dimension is the model analysis reports as presented in Table 4. The 2<sup>nd</sup> dimension is the prototype model with the highest efficiency and highest accuracy as shown in Table 5.

**Table 4.** Decision tree classification model analysis

Depth_DT	k-Fold	Accuracy	Precision	Recall	AUC
2*	10	76.04%	75.36%	91.87%	79.20%
	35	77.16%	77.70%	89.10%	83.10%
	50	77.83%	79.06%	89.17%	84.70%
	100	78.25%	70.83%	88.83%	84.50%
	leave-one-out*	78.30%*	77.32%	88.89%	49.30%
3	10	75.54%	76.46%	89.28%	78.90%
	35	77.66%	78.37%	88.21%	84.30%
	50	77.36%	79.53%	87.63%	84.70%
	100	77.75%	80.08%	87.17%	85.00%
	leave-one-out	77.81%	77.78%	86.75%	49.50%

(Continued)

**Table 4.** Decision tree classification model analysis (Continued)

Depth_DT	k-Fold	Accuracy	Precision	Recall	AUC
4	10	75.54%	76.87%	88.84%	79.10%
	35	77.66%	78.37%	88.21%	84.30%
	50	77.11%	79.47%	87.23%	84.60%
	100	77.50%	79.92%	86.67%	84.50%
	leave-one-out	77.31%	77.61%	85.90%	44.50%

Note: Depth\_DT = depth of decision tree model.

Table 4 shows the analysis of Decision Tree modeling. It found that the Decision Tree model at a depth of level 2 and a test of the most effective of the leave-one-out cross-validation method is equal to 78.30% of the highest accuracy. The test results are shown in Table 5.

**Table 5.** Highest decision tree model performance

Accuracy: 78.30%		Actual Class		Class Precision
		True Dropped	True Graduated	
Prediction Class	Pred. Dropped	106	26	80.30%
	Pred. Graduated	61	208	77.32%
	Class Recall	63.47%	88.89%	

Table 5 presents the details of the most effective Decision Tree models. The details are from Table 4, where it was found that the Decision Tree model with depth 2 and tested with cross-validation method by leave-one-out was the most effective. It has the accuracy rate equal to 78.30%, the precision rate equal to 77.32%, the recall rate equal to 88.89%, and the AUC rate equal to 49.39%. It can be concluded that this model is suitable for this approach.

**Results of the forward selection model.** The analysis results of the Forward Selection model were shown in two dimensions, much like the analysis of a Decision Tree model to compare reasonable results. The 1<sup>st</sup> dimension is the model analysis reports as presented in Table 6. The 2<sup>nd</sup> dimension is the prototype model with the highest efficiency and highest accuracy as shown in Table 7.

**Table 6.** Forward selection model analysis

Num_Att	Depth_DT	k-Fold	Accuracy	Precision	Recall	AUC
2	3	100	81.75%	85.15%	82.67%	87.62%
3	4	10	82.05%	82.45%	88.80%	87.50%
4*	4*	leave-one-out*	82.29%*	85.59%*	83.76%*	45.14%
5	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
6	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
7	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
8	4	leave-one-out	82.29%	85.59%	83.76%	45.14%

(Continued)

**Table 6.** Forward selection model analysis (*Continued*)

Num_Att	Depth_DT	k-Fold	Accuracy	Precision	Recall	AUC
9	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
10	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
11	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
12	4	leave-one-out	82.29%	85.59%	83.76%	45.14%
13	4	leave-one-out	82.29%	85.59%	83.76%	45.14%

Notes: Num\_Att = number of attributes, Depth\_DT = depth of decision tree model.

Table 6 shows the analysis of Forward Selection modeling. It found that the Forward Selection model with 4 attributes at a depth of level 4 and a test of the most effective of the leave-one-out cross-validation method is equal to 82.29% of the highest accuracy. The test results are shown in Table 7. Please note that Table 6 presents only the data with the highest accuracy in each of the increased attributes, with complete analysis provided in the link: <https://bit.ly/3xVYe8s>.

The attributes that are significant to the model analysis consist of four courses: 001103 Thai Language Skills, 221100 Business Mathematics, 221110 Fundamental Information Technology, and 221120 Introduction to Programming.

**Table 7.** Highest decision tree model performance

Accuracy: 82.29%		Actual Class		Class Precision
		<i>True Dropped</i>	<i>True Graduated</i>	
Prediction Class	<i>Pred. Dropped</i>	134	38	77.91%
	<i>Pred. Graduated</i>	33	196	85.59%
	<i>Class Recall</i>	80.24%	83.76%	

Table 7 presents the details of the most effective decision tree models. The details are from Table 6, where it was found that the Forward Selection model with 4 attributes at depth 4 and tested with cross-validation method by leave-one-out was the most effective. It has the accuracy rate equal to 82.29%, the precision rate equal to 85.59%, the recall rate equal to 83.76%, and the AUC rate equal to 45.14%. It can be concluded that this model is suitable for this approach.

**Results of the backward elimination model.** The analysis results of the Backward Elimination model were displayed in two dimensions, much like the analysis of a Decision Tree model to compare reasonable results. The 1<sup>st</sup> dimension is the model analysis reports as presented in Table 8. The 2<sup>nd</sup> dimension is the prototype model with the highest efficiency and highest accuracy as shown in Table 9.

**Table 8.** Backward elimination model analysis

Num_Att	Depth_DT	k-Fold	Accuracy	Precision	Recall	AUC
13	4	leave-one-out	82.29%	87.21%	81.62%	42.27%
12	4	leave-one-out	82.29%	87.21%	81.62%	42.27%
11	4	leave-one-out	82.29%	87.21%	81.62%	42.27%
10	4	leave-one-out	82.29%	87.21%	81.62%	42.27%
9	4	leave-one-out	82.29%	87.21%	81.62%	42.27%
8*	4*	leave-one-out*	82.29%*	87.21%*	81.62%*	42.27%
7	3	10	82.05%	84.94%	85.04%	88.59%
6	4	leave-one-out	81.50%	87.21%	81.62%	42.27%
5	4	10	82.05%	83.23%	87.64%	87.50%
4	4	50	81.78%	87.20%	83.17%	88.64%
3	4	20	80.52%	83.36%	85.87%	86.00%
2	3	10	82.05%	82.27%	88.42%	86.77%

Notes: Num\_Att = number of attributes, Depth\_DT = depth of decision tree model.

Table 8 shows the analysis of Backward Elimination modeling. It found that the Backward Elimination model with 8 attributes at a depth of level 4 and a test of the most effective of the leave-one-out cross-validation method is equal to 82.29% of the highest accuracy. The test results are shown in Table 9. Please note that Table 8 presents only the data with the highest accuracy in each of the reduced attributes, with complete analysis provided in the link: <https://bit.ly/3xVYe8s>.

The attributes that are significant to the model analysis consist of eight courses: 001103 Thai Language Skills, 001112 Developmental English, 003134 Civilization and Indigenous Wisdom, 126100 Introduction to Economics, 128221 Principles of Marketing, 221100 Business Mathematics, 221110 Fundamental Information Technology, and 221120 Introduction to Programming.

**Table 9.** Highest decision tree model performance

Accuracy: 82.29%		Actual Class		Class Precision
		<i>True Dropped</i>	<i>True Graduated</i>	
Prediction Class	<i>Pred. Dropped</i>	139	43	76.37%
	<i>Pred. Graduated</i>	28	191	87.21%
	<i>Class Recall</i>	83.23%	81.62%	

Table 9 presents the details of the most effective Backward Elimination models. The details are from Table 8, where it was found that the Backward Elimination model with 8 attributes, depth 4 and tested with cross-validation method by leave-one-out was the most effective. It has the accuracy rate equal to 82.29%, the precision rate equal to 87.21%, the recall rate equal to 81.62%, and the AUC rate equal to 42.27%. It can be concluded that this model is suitable for this approach.

*Notice:* From the analysis of the three models. It discovered that there are two models with the highest accuracy, Forward Selection and Backward Elimination methods, with the same accuracy rate at 82.29%. It is imperative to select only one model, which the researcher decides to use for Forward Selection model, as the recall rates (83.76% per 81.62%) and AUC rates (50.00% per 42.27%) are considered higher. The researchers show the rule model in Table 10.

**Table 10.** Rule model

<b>Rule Model</b>	<b>Dropped</b>	<b>Graduated</b>
if 001103 = B then Graduated	1 (6.46%)	12 (77.58%)
if 001103 = B+ then Graduated	0 (0.00%)	5 (84.04%)
if 001103 = C and 221100 = A then Graduated	2 (10.51%)	14 (73.54%)
if 001103 = C and 221100 = B then Graduated	0 (0.00%)	6 (84.04%)
if 001103 = C and 221100 = B+ then Graduated	0 (0.00%)	7 (84.04%)
if 001103 = C and 221100 = C then Graduated	0 (0.00%)	13 (84.04%)
if 001103 = C and 221100 = C+ then Graduated	1 (8.40%)	9 (75.64%)
if 001103 = C and 221100 = D then Graduated	1 (9.34%)	8 (74.70%)
if 001103 = C and 221100 = D+ then Graduated	4 (22.41%)	11 (61.63%)
if 001103 = C and 221100 = F then Dropped	10 (56.03%)	5 (28.01%)
if 001103 = C and 221100 = W then Graduated	4 (13.45%)	21 (70.59%)
if 001103 = C+ then Graduated	2 (3.06%)	53 (80.98%)
if 001103 = D then Dropped	32 (49.80%)	22 (34.24%)
if 001103 = D+ and 221100 = A then Graduated	0 (0.00%)	2 (84.04%)
if 001103 = D+ and 221100 = B then Graduated	0 (0.00%)	2 (84.04%)
if 001103 = D+ and 221100 = C then Graduated	1 (9.34%)	8 (74.70%)
if 001103 = D+ and 221100 = C+ then Graduated	1 (28.01%)	2 (56.03%)
if 001103 = D+ and 221100 = D then Graduated	4 (33.62%)	6 (50.42%)
if 001103 = D+ and 221100 = D+ then Graduated	1 (16.81%)	4 (67.23%)
if 001103 = D+ and 221100 = F then Graduated	9 (37.82%)	11 (46.22%)
if 001103 = D+ and 221100 = W and 221120 = A then Graduated	0 (0.00%)	2 (84.04%)
if 001103 = D+ and 221100 = W and 221120 = C then Graduated	0 (0.00%)	3 (84.04%)
if 001103 = D+ and 221100 = W and 221120 = C+ then Graduated	0 (0.00%)	2 (84.04%)
if 001103 = D+ and 221100 = W and 221120 = D then Dropped	2 (84.04%)	0 (0.00%)
if 001103 = D+ and 221100 = W and 221120 = D+ then Dropped	4 (84.04%)	0 (0.00%)
if 001103 = D+ and 221100 = W and 221120 = F then Dropped	3 (63.03%)	1 (21.01%)
if 001103 = D+ and 221100 = W and 221120 = W then Dropped	9 (75.64%)	1 (8.40%)
if 001103 = F then Dropped	41 (76.57%)	4 (7.47%)
if 001103 = W then Dropped	35 (84.04%)	0 (0.00%)
<i>Correct: 337 out of 401 training examples. (84.04%)</i>		

Table 10 shows the Rule Model. It discovered that the model generated when tested with the collected data had highly accurate predictions, with 337 accurate predictions out of 401 (equal to 84.04 percent). This Rule Model was developed as an application which is presented in the next section.

### 3.3 Prototype of the prediction application model

This section presents the user interface of the application at <https://bit.ly/3yiP8Tk>. It consists of five main pages, as presented in Figures 7 to Figures 11.

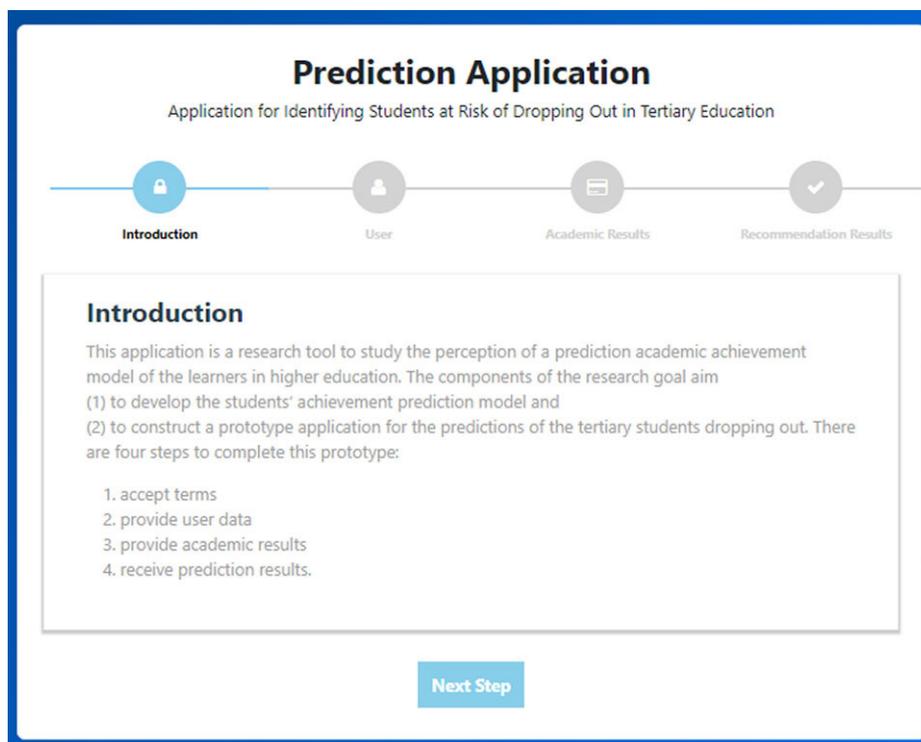


Fig. 7. The user interface of the application

Figure 7 is the initial process of the application. It is an introduction page with the application name and an agreement. When the user agrees to the agreement, they will go to the next page.

**Prediction Application**  
Application for Identifying Students at Risk of Dropping Out in Tertiary Education

Introduction User Academic Results Recommendation Results

**Personal Information**

First Name  
Last Name  
Student ID  
Email  
Phone

Previous Next Step

Fig. 8. The user interface of the application

Figure 8 shows the general user information for the application. It consists of first name, last name, email address and phone number. The purpose of this section is to communicate with future users.

**Prediction Application**  
Application for Identifying Students at Risk of Dropping Out in Tertiary Education

Introduction User Academic Results Recommendation Results

**Grade Information**

001103 Thai Language Skills  
A B+ B C+ C D+ D F W

221100 Business Mathematics  
A B+ B C+ C D+ D F W

221120 Introduction to Programming  
A B+ B C+ C D+ D F W

221110 Fundamental Information Technology  
A B+ B C+ C D+ D F W

Previous Confirm

Fig. 9. The user interface of the application

Figure 9 shows the filling of academic results in the application to predict the students' academic achievement. The data in this page contains the academic results in four courses: 001103 Thai Language Skills, 221100 Business Mathematics, 221110 Fundamental Information Technology, and 221120 Introduction to Programming.

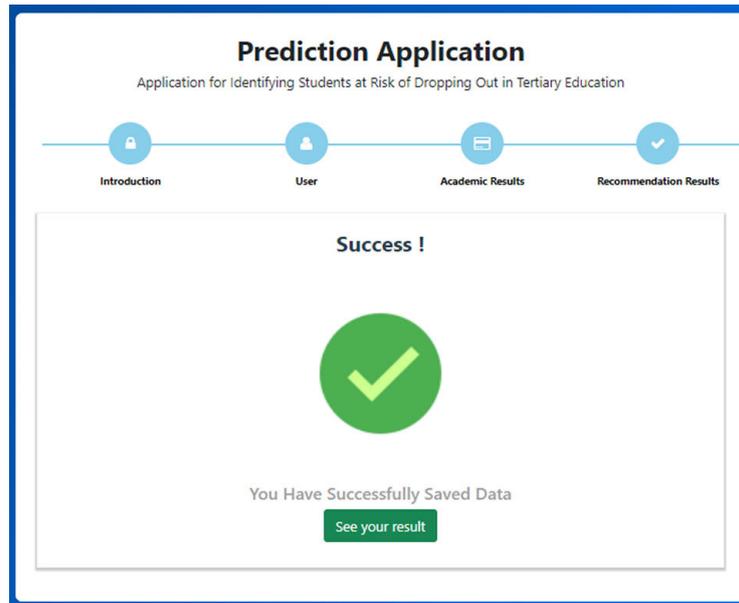


Fig. 10. The user interface of the application

Figure 10 shows the complete user academic achievement results that is calculated and displayed in Figure 11.

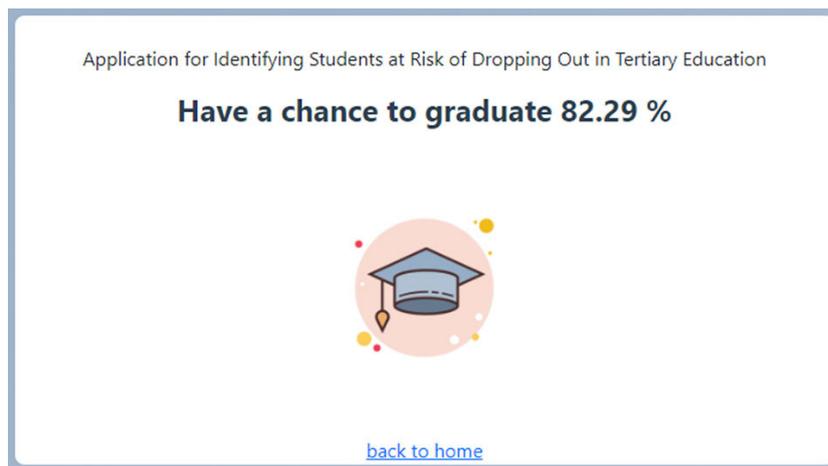


Fig. 11. The user interface of the application

Figure 11 shows the analysis results according to the model obtained from Table 10. The result page is to show the percentage of the opportunity that a student will graduate or dropping out.

## **4 Research discussions**

The discussion was addressed on two major areas, aligned with the research objective: The student achievement prediction model and the application for tertiary student dropout prediction.

### **4.1 Model discussions**

Model analysis issues in this research, the researcher presented the three important models. It consists of model from Decision Trees Classification technique, model from Forward Selection techniques, and model from Backward Elimination technique.

The researchers discovered that the model from the Decision Tree Classification technique with the highest accuracy was the model with depth at level 2 and tested with a leave-one-out Cross-Validation method had the highest accuracy of 78.30%. In addition, the model has a precision of 77.32%, a recall of 88.89%, and AUC rate of 49.30%. While the model from the Forward Selection technique with the highest accuracy was the model with 4 attributes, depth at level 4 and tested with a leave-one-out Cross-Validation method had the highest accuracy of 82.29%. It also has a precision of 85.59%, a recall of 83.76%, and AUC rate of 45.14%. Lastly, the model from the Backward Elimination technique with the highest accuracy was the model with 8 attributes, depth at level 4 and tested with a leave-one-out Cross-Validation method had the highest accuracy of 82.29%. It also has a precision of 82.29%, a recall of 87.21%, and AUC rate of 42.27%.

By analyzing the three models, the researchers concluded that a reasonable and appropriate model to be programmed is the Forward Selection model. The rationale is determined by indicators including Accuracy, Precision, Recall, and AUC. The proof is that the Forward Selection Model has both a significantly higher Recall rate (83.76% per 81.62%) and AUC rate (50.00% per 42.27%).

### **4.2 Application discussions**

For discussion of the application, the researcher used a questionnaire to assess their satisfaction with the application. The respondents were 26 students from the Business Computer program in the 2<sup>nd</sup> semester of the academic year 2020 at the School of Information and Communication Technology, the University of Phayao.

Software Testing is the process of evaluating and improving the quality of software by identifying software errors or bugs. It can identify useful approaches for improving and modifying the software. The method used is the Black-Box testing. Black-Box testing is a software testing method that validates an application's functionality without having to understand its structure or functionality within the software. This testing method can be used for all levels of software testing. The 1<sup>st</sup> topic is the usability test

(U-Test). It is an evaluation of a program’s ability to interact with the user of an application. The 2<sup>nd</sup> topic is the function test (F-Test). It is an assessment of the correctness of the application operation.

The 3<sup>rd</sup> topic is the functional requirement test (FR-Test). It is an assessment of the application’s capabilities according to the needs of the users. The last topic is the security test (SC-Test). It is intended to evaluate an application in the field of security and data protection.

For the level of satisfaction of the application, questionnaires were characterized by using a Likert Scale. It has set a rating to measure opinions into 5 levels: Satisfaction level 5 means strongly agree, satisfaction level 4 means agree, satisfaction level 3 means neither agree nor disagree, satisfaction level 2 means disagree, and satisfaction level 1 means strongly disagree. The acceptance criteria used in the interpretation of satisfaction levels. It calculates the mean which is within 5 specified levels: The average score equal to 4.21–5.00 is the highest accepted level, the average score equal to 3.41–4.20 is the high accepted level, the average score equal to 2.61–3.40 is the accepted level, the average score equal to 1.81–2.60 is the low accepted level, and the average score equal to 1.00–1.80 is the lowest accepted level. The report analyzes the attitudes and satisfaction with the application are presented in Table 11.

**Table 11.** The attitude and satisfaction toward the application

Questions	Mean	S.D.	Interpretation
<b><i>The Usability Test (U-Test)</i></b>			
Language and Symbols convey meaning clearly	3.53	1.12	High Accepted
Layout management is easy to understand.	3.86	0.92	High Accepted
Text, Font, Colors, and Images are appropriate.	3.87	0.87	High Accepted
Visibility of the information is standardized.	4.16	0.82	High Accepted
Average:	3.86	0.93	High Accepted
<b><i>The Function Test (F-Test)</i></b>			
Appropriateness of the acting results	4.00	0.82	High Accepted
Appropriateness of the advice from analysis	4.15	0.78	High Accepted
Appropriateness of the analysis presentation	4.15	0.85	High Accepted
Appropriateness of the results analysis	3.98	0.93	High Accepted
Average:	4.07	0.85	High Accepted
<b><i>The Functional Requirement Test (FR-Test)</i></b>			
Ability to control program execution	3.98	0.93	High Accepted
Presenting information according to the needs	3.82	0.89	High Accepted
User response time	3.84	0.89	High Accepted
Average:	3.88	0.90	High Accepted
<b><i>The Security Test (SC-Test)</i></b>			
Correctness of the application in authentication	3.84	0.89	High Accepted
User confidence in the application	3.88	0.9	High Accepted
Average:	3.86	0.90	High Accepted
<b>Total Average:</b>	<b>3.93</b>	<b>0.89</b>	<b>High Accepted</b>

Table 11 shows the attitude and satisfaction toward the application for identifying students at risk of dropping out in tertiary education. Overall satisfaction and attitude, it was found that respondents were highly satisfied with the application (mean = 3.93, S.D. = 0.89). The most recognized aspect is the function test (F-Test) with mean equal to 4.07 (S.D. = 0.85). The second most recognized ranking is the functional requirement test (FR-Test) with mean equal to 3.88 (S.D. = 0.90). The third recognized ranking is the usability test (U-Test) with mean equal to 3.86 (S.D. = 0.93). Finally, the last recognized ranking is the security test (SC-Test) with mean equal to 3.86 (S.D. = 0.90).

In summary, the applications being developed have a high level of acceptance that researchers will implement in the future.

## **5 Conclusion**

The situation affecting the Thai education process is the inability to predict the learners' academic achievement. Thus, this research objective is to identify the risk of dropping out in tertiary students with an application. The research goals are (1) to develop the students' achievement prediction model, and (2) to construct a prototype application for the predictions of the tertiary students dropping out.

The research tools consisted of three parts, (1) tool for developing predictive prototypes uses a tool called the CRISP-DM process with Decision Tree Classification, Feature Selection methods, Confusion Matrix performance, Cross-Validation methods, Accuracy, Precision and Recall measurements, (2) tool for application development used the SDLC with V-method, and (3) tool to assess application satisfaction used questionnaires and statistical analysis. Data sample were collected from 401 students enrolled in the Business Computer Program at the School of Information and Communication Technology, University of Phayao during the academic year 2012–2016. The results showed that the prediction model had a very high percentage of accuracy (82.29%). The prototype test results with the data gathered had a very high score level (84.04%; correct 337 out of 401 training examples). An overview of the underlying application with the utmost integrity by the researchers planned to put the application to the test in the first semester of the academic year 2021 at the School of Information Technology and Communication, University of Phayao.

For future research, the researchers plan to conduct a research testing with students in Computer Business, the School of Information and Communication Technology, University of Phayao in the following academic year.

## **6 Future works**

For future research, the researchers plan to conduct a research testing with students in Computer Business, the School of Information and Communication Technology, University of Phayao in the next academic year. In addition, the researchers plan to create a mobile application for mentors in the University of Phayao to monitor learner on both Android and iOS systems.

## 7 Conflict of interest and research ethics

The researchers declare no conflict of interest. Research ethics: the researchers is allowed to conduct this research according to the announcement of the University of Phayao: No. 2/020/63 on April 22, 2020.

## 8 Acknowledgment

This research was supported by three universities: (1) the School of Information and Communication Technology, University of Phayao, (2) the Faculty of Science and Technology, Rajabhat Maha Sarakham University, and (3) the Faculty of Information Technology, Rajabhat Maha Sarakham University. The researchers would like to thank the supervisor (Associate Professor Dr. Sittichai Bussaman), all advisors, all instructors, all students, all staffs, and all respondents for their entire support. Finally, this research would like to thank for the grant by the project of the Unit of Excellence in Applied Informatics for Modernization (FF64-UoE004) from the University of Phayao.

## 9 References

- [1] 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, doi: <https://doi.org/10.1504/IJMLO.2019.096469>
- [2] N. Mileva, "The Effectiveness of Mobile Learning in the Form of Performance Support System in Higher Education," *Int. J. Interact. Mob. Technol. IJIM*, vol. 5, no. 4, Art. no. 4, Oct. 2011, doi: <https://doi.org/10.3991/ijim.v5i4.1692>
- [3] 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. IJET*, vol. 14, no. 21, pp. 205–226, Nov. 2019, doi: <https://doi.org/10.3991/ijet.v14i21.11034>
- [4] C. N. Udanor and O. U. Oparaku, "A Performance Evaluation of a Multi-Agent Mobile Learning System," *Int. J. Interact. Mob. Technol. IJIM*, vol. 10, no. 2, Art. no. 2, Apr. 2016, doi: <https://doi.org/10.3991/ijim.v10i2.4873>
- [5] A. Nousseir and Y. Fathy, "A Mobile Application for Early Prediction of Student Performance Using Fuzzy Logic and Artificial Neural Networks," *Int. J. Interact. Mob. Technol. IJIM*, vol. 14, no. 02, Art. no. 02, Feb. 2020, doi: <https://doi.org/10.3991/ijim.v14i02.10940>
- [6] H. Elazhary, "Context-aware Cloud-based Mobile Application for Assessment and Training of Visual Cognitive Abilities," *Int. J. Interact. Mob. Technol. IJIM*, vol. 11, no. 6, Art. no. 6, Nov. 2017, doi: <https://doi.org/10.3991/ijim.v11i6.7438>
- [7] K. Y. S. Putri, Z. B. Abdullah, E. Nugrahaeni, R. Darmawan, and L. Latifa, "Learning Management Strategy of Communication Studies through Blended Learning in Higher Education," *Int. J. Interact. Mob. Technol. IJIM*, vol. 14, no. 16, Art. no. 16, Sep. 2020, doi: <https://doi.org/10.3991/ijim.v14i16.15725>
- [8] 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, doi: [https://doi.org/10.1016/S1056-4993\(18\)30106-8](https://doi.org/10.1016/S1056-4993(18)30106-8)
- [9] Y. Hu *et al.*, "Investigating the Attentional Bias and Information Processing Mechanism of Mobile Phone Aaddicts towards Emotional Information," *Comput. Hum. Behav.*, vol. 110, p. 106378, Sep. 2020, doi: <https://doi.org/10.1016/j.chb.2020.106378>

- [10] F.-Y. Hong, S.-I. Chiu, and D.-H. Huang, “A Model of the Relationship between Psychological Characteristics, Mobile Phone Addiction and Use of Mobile Phones by Taiwanese University Female Students,” *Comput. Hum. Behav.*, vol. 28, no. 6, pp. 2152–2159, Nov. 2012, doi: <https://doi.org/10.1016/j.chb.2012.06.020>
- [11] 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.
- [12] 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, doi: <https://doi.org/10.3991/ijet.v14i19.11177>
- [13] A. Jovic, K. Brkic, and N. Bogunovic, “A Review of Feature Selection Methods with Applications,” in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1200–1205, May 2015, doi: <https://doi.org/10.1109/MIPRO.2015.7160458>
- [14] G. Chen and J. Chen, “A Novel Wrapper Method for Feature Selection and its Applications,” *Neurocomputing*, vol. 159, pp. 219–226, Jul. 2015, doi: <https://doi.org/10.1016/j.neucom.2015.01.070>
- [15] M. McHugh, O. Cawley, F. McCaffry, I. Richardson, and X. Wang, “An agile V-model for medical device software development to overcome the challenges with plan-driven software development lifecycles,” in *2013 5th International Workshop on Software Engineering in Health Care (SEHC)*, pp. 12–19, May 2013, doi: <https://doi.org/10.1109/SEHC.2013.6602471>
- [16] S. Balaji and M. S. Murugaiyan, “Waterfall vs. V-Model vs. Agile: A comparative study on SDLC,” *Int. J. Inf. Technol. Bus. Manag.*, vol. 2, no. 1, pp. 26–30, 2012.

## 10 Authors

**Pratya Nuankaew** received the Bachelor of Education (Educational Technology) in 2001, the Master of Science (Information Technology) in 2008 from Naresuan University, and Doctor of Philosophy (Computer Engineering) in 2018 from Mae Fah Luang University. He is currently an instructor at the School of Information and Communication Technology, University of Phayao, Phayao, 56000, Thailand. His research interests are in applied informatics technology, educational data mining, educational engineering, educational technology, informatics, applications, learning analytics, learning strategies for lifelong learning, online mentoring model, social network analysis, ubiquitous computing, and context aware.

**Patchara Nasa-Ngium** received the B.S. degree in computer science from Rajabhat Maha Sarakham University, the M.S. degree and Ph.D. degree in computer science from Khon Kaen University, Thailand. He is also a Lecturer with the Department of Computer Science, Rajabhat Maha Sarakham University. His research interests include artificial intelligence, machine learning, evolutionary computing, and data mining.

**Wongpanya Nuankaew** received a B.Sc. degree in Computer Science in 2004, and M.Sc. degree in Information Technology in 2007 from Naresuan University. She is currently a lecturer at the Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand.

Article submitted 2021-05-18. Resubmitted 2021-06-28. Final acceptance 2021-07-03. Final version published as submitted by the authors.