

Prediction Model of Student Achievement in Business Computer Disciplines

Learning Strategies for Lifelong Learning

<https://doi.org/10.3991/ijet.v15i20.15273>

Pratya Nuankaew ^(✉)

University of Phayao, Mueang Phayao, Thailand
pratya.nu@up.ac.th

Wongpanya Nuankaew

Rajabhat Maha Sarakham University, Maha Sarakham, Thailand

Direk Teeraputon

Naresuan University, Mueang Phitsanulok, Thailand

Kanakarn Phanniphong

Rajamangala University of Technology Tawan-Ok, Si Racha, Thailand

Sittichai Bussaman

Rajabhat Maha Sarakham University, Maha Sarakham, Thailand

Abstract—An educational program that does not accept the change of disruptive technology will inevitably result in future destruction. There are two objectives including (1) to construct reasonable students' dropout prediction model for business computer disciplines, and (2) to evaluate the model performance. Data collected consists of 2,017 records from students who enrolled in the business computer program at the School of Information and Communication Technology, the University of Phayao. Research tools are divided into two parts. (1) Modelling; it consisted of the Artificial Neural Network Algorithm, Decision Tree Algorithm, and Naïve Bayes Algorithm. (2) Model testing; it consisted of the confusion matrix performance, accuracy, precision, and recall measurement. It is a clear innovation in the research that the researcher combines the knowledge of data science in analysis to improve the academic achievement of students in higher education in Thailand. From the analysis results, it shows that the model developed from using Artificial Neural Network algorithms has the highest accuracy in the first three data sets (89.04%, 92.70% and 93.71%), and the last model is appropriate for Naïve Bayes algorithm (91.68%). Finally, it is necessary to conduct additional research and present research results to relevant parties and organizations.

Keywords—Prediction Model, Students Academic Achievement, Educational Data Mining, Learning Analytics.

1 Introduction

In the field of education, there is a significant need for change which is called “Disruptive Education Technologies” [1], [2]. This innovation is most likely to revolutionize the learning structure. It will not only change the educational content and experience of students but will influence students’ learning interest and retention [3]. What is about to change the world of education is digital technology that will disrupt higher learning [4]. Therefore, using the right innovation will support and create benefits for the users.

From this situation, the university's education program has also been affected. According to UNESCO’s 2017/8 Global Education Monitoring (GEM) Report, there are still many challenges for raising the quality of education in Thailand: The number of students who have completed primary education is 99%, but only 85% have completed a lower secondary education. At the lower secondary level, only 50% of the students have passed the reading exam, and only 46% of the students have passed the mathematics exam. Moreover, there are 3.9 million adults who are unable to read a simple sentence [5]. This report is the starting point of our research, in which the researchers found that the preliminary quality of students in the university is diverse. Some students have the potential and readiness for academic learning. On the other hand, some students do not have easy access to knowledge. Preliminary results are a problem of students’ academic performance at the university level, which results in students having a low performance leading them to drop out of the program [6], [7].

This research aims to present and reflect the problems about student achievement and dropout of university students by limiting the study scope to only one program at the University of Phayao. The expected results are aimed to know the causes and factors that reflect the problem of students’ dropout in the Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao and bring this research to further development as shown in the conceptual framework in Figure 1.

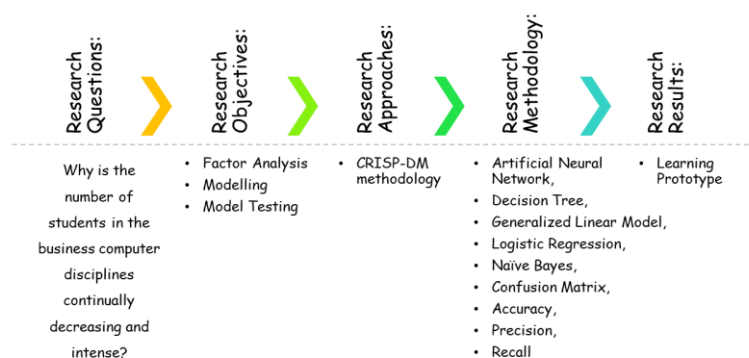


Fig. 1. Conceptual Framework

1.1 Research objectives

There are two main objectives on this paper: (1) to construct reasonable students' dropout prediction model for Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao, and (2) to evaluate the performance and select the best predictions of the dropout model of students who did not gain academic achievement for the Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao.

1.2 Research approaches

The researchers approached the work with a design according to the data mining principle called CRISP-DM [8]. It consists of six steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The data collected and accessed for study in the research are the data of 2,017 business computer students from the academic year 2001-2019 at the School of Information and Communication Technology, at the University of Phayao, Thailand.

Research tools are divided into two main parts. The first part is the development of predictive models [9]–[12], it consisted of the Artificial Neural Network Algorithm (A-NN), Decision Tree Algorithm (DT), and Naïve Bayes Algorithm (NB). The second part is the model performance testing as described in the evaluation section, it consisted of the confusion matrix performance, accuracy, precision, and recall measurement [10].

2 Literature Reviews and Related Works

2.1 Educational system in the university

The education system in the university is in accordance with the National Education Act of Thailand which includes the National Education Act B.E. 2542 (1999) and Amendments of the Second National Education Act B.E. 2545 (2002) [13], [14], and The National Qualifications Framework for Higher Education in Thailand [15].

The National Education Act requires three types of education: Formal education specifies the aims, methods, curricula, duration, assessment, and evaluation that are conditional to its completion. Non-formal education has flexibility in determining the aims, modalities, management procedures, duration, assessment and evaluation conditional to its completion. The contents and curricula for non-formal education shall be appropriate, responding to the requirements, and meeting the needs of individual groups of learners. Finally, informal education enables learners to learn by themselves according to their interests, potentialities, readiness and opportunities available from individuals, society, environment, media, or other sources of knowledge.

In addition, formal education is divided into two levels: basic education and higher education. The formal education structure in Thailand is shown in Figure 2.

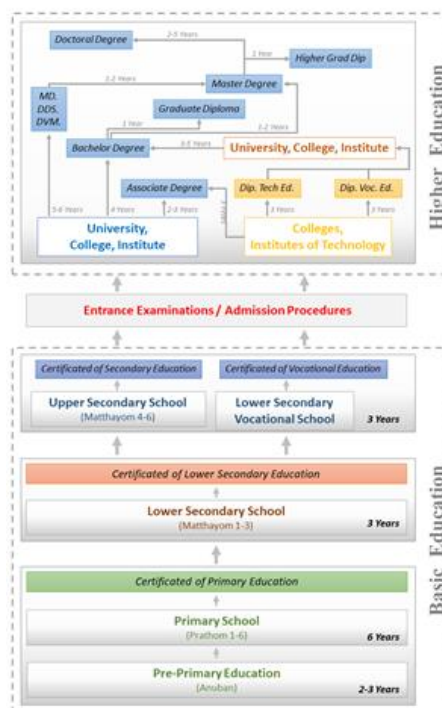


Fig. 2. Formal education structure in Thailand (<https://bit.ly/2VZTWuA>)

Figure 2 shows the structure and duration of the Thai education system, which takes an average of almost two decades. In addition, the university education system must refer to the National Qualifications Framework for Higher Education in Thailand [13]. It has been designed and defined with many qualifications and standards for getting students to complete their studies within a specific time frame. For example, there are five characteristics of programs and expected outcomes in the domains of learning: ethical and moral development, knowledge, cognitive skills, interpersonal skills and responsibility, and analytical and communication skills. Moreover, the Thai Qualifications Framework for Higher Education (TQF: HEd), which is modified from The National Qualifications Framework, has specified too many requirements, according to numerous research reports [10-12].

Therefore, it can be concluded that the Thai education system was not truly designed in line with the learner's context.

2.2 University's competitiveness and dropout's impact

Currently, higher education in Thailand consists of 83 public higher education institutions, consisting of 26 autonomous universities, 10 public universities, 38 Rajabhat Universities (developed from teacher training colleges but now offering a broad range of programs beyond focusing on developing local communities), and 9

Rajamangala Universities of Technology (developed from institutes of technology, and focusing on science and technology). In addition, there are 72 private higher education institutions.

There are 155 universities and more than 59 campuses in Thailand. The number of students enrolled in the university has decreased steadily as shown in Table 2. The explanation of the decline comes from the competition of new students trying to get a place in the university of their choice. In 2017, two educational institutions in Thailand closed down, namely Asian University and Srisophon College. While the number of students in higher education institutions that entered the university during the year 2013-2017 is shown in Table 1.

Table 1. Number of students in higher education institutions during the year 2013-2017

Education Level	Academic Years				
	2017	2016	2015	2014	2013
Associate degree	329,673	348,060	344,377	312,770	316,288
Bachelor's degree	1,715,976	1,780,382	1,851,653	1,843,477	1,881,816
Master's degree	143,961	156,347	188,471	193,679	207,145
Doctoral degree	24,472	23,995	24,742	24,487	25,394
Total:	2,214,082	2,308,784	2,409,243	2,374,413	2,430,643

Table 1 shows the number of students decreasing continuously. In 2017, the total number of students decreased by 216,561 students. In addition, the bachelor's degree level has decreased significantly with the number of students reduced to 165,840 students. Moreover, the Ministry of Education in Thailand has reported a decrease in the number of students enrolled in higher education in 2016. There are 81.26 percent enrolled in the first year of higher education, or 541,720 students. From the situation, educational institutions are needed more and the population has been continuously decreasing, thus concluding that the competition for the learners will be more intense in the future.

The Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao has clearly felt the impact [10], [12]. The institution has discussed the statistics of students continuing to decline in this research. This research is needed to solve the dropout problem of university students. Which, the researcher therefore pays special attention to this problem.

3 Research Methodology

This research methodology was designed and implemented in accordance with the CRISP-DM methodology [8]. It has six phases to follow: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

3.1 Business understanding

This phase focuses on questioning and studying the needs of the Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao from the past twenty years. The objective is to use research to find answers that have been developed as a model for problem solving, the number of students in the business computer disciplines, and the dropout problem of students. How should it be solved?

In the past, the state of the business computer disciplines used to have a large number of students but it has continued to decrease over the past twenty years, while the higher dropout rate makes it a very serious threat to the discipline's existence. This is converted to the viewpoint of the success of an organization which is an educational institution that wants to promote and support human resource development which is the country's strength.

3.2 Data understanding

The data understanding phase aims to define the scope of data collection in response to the research questions. Normally, there are four-steps which includes collecting the initial data, describing data, exploring data, and verifying data quality [8].

The purpose of this phase is to gather a comprehensive and complete information of the Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao, Thailand. It consists of statistics of students enrolled in the program during the twenty years (academic year 2001-2019), dropout statistics of students in the program during the twenty years (academic year 2001-2019), current student statistics, and details of individual study results including grade point average (GPA), registration history for each semester, academic results in each course, student status, and other related information.

Figure 3 below displays a diagram showing the process of data collection, the process of analyzing and determining the scope of the data, and the process of planning the data management. Moreover, Figure 4 shows the process of obtaining data from the university.

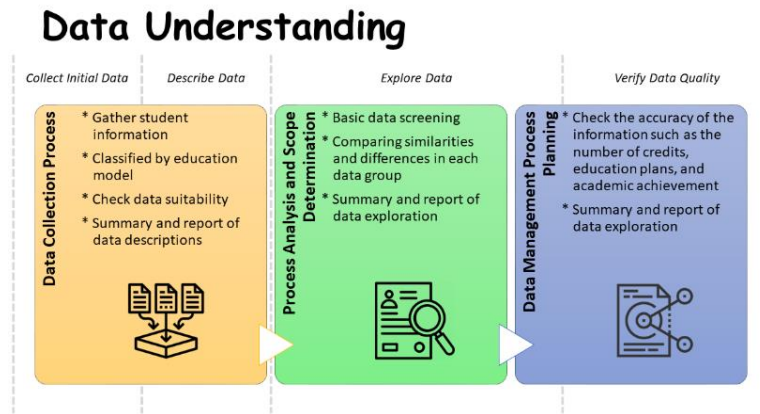


Fig. 3. Data scoping and Data collection diagram

Figure 3 shows the working process of the data understanding phase. It shows the relationship between CRISP-DM [8] theory and research operations. It can be seen that this phase takes a lot of time due to the fact that there are sub-processes and details that need to be handled to obtain quality data.

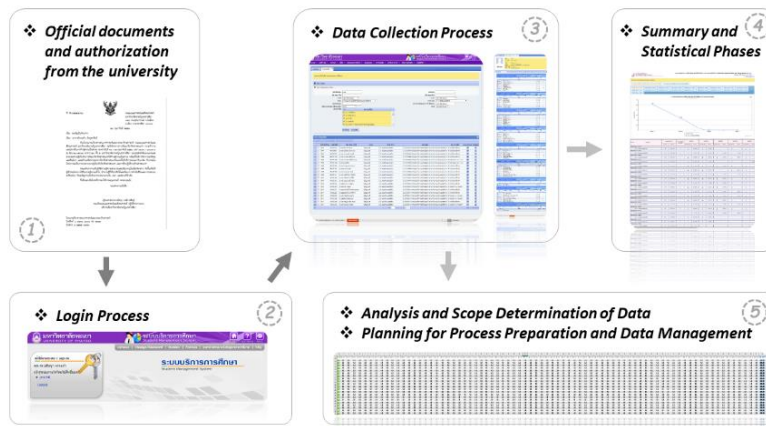


Fig. 4. Data management processes

Figure 4 shows the data acquisition procedure by referring to the procedures in the research methodology. The steps shown in Figure 4 consist of five steps: 1) requesting permission to access and use student data from the university, 2) login process, 3) student data collection, 4) statistical summary, and 5) prepare data for developing models. The preparation and distribution of the data are shown in section of data preparation and the selection of tools for developing the model are shown in the modeling section.

3.3 Data preparation

The data preparation phase covers all activities to complete the final data collection. It is intended to prepare data that will be fed into the modeling tool from the initial raw data. It consists of five main steps which include selecting data, cleaning data, constructing data, integrating data, and formalizing data [8].

The data collected were 2,017 students in the Business Computer Disciplines at the School of Information and Communication Technology, at the University of Phayao, Thailand. The data is divided into five sets from the scope of research shown in Table 2 to Table 4, respectively.

Notation, the scope of the research is divided according to the revised curriculum, which will be updated every three to five years.

Table 2. Data collected

Types and Data Series	Data Set No. 1	Data Set No. 2	Data Set No. 3	Data Set No. 4	Data Set No. 5	Total
Graduation (scheduled)	286 (14.18%)	356 (17.65%)	276 (13.68%)	127 (6.30%)	0	1,045 (51.81%)
Graduation (delayed)	29 (1.44%)	36 (1.78%)	60 (2.97%)	57 (2.83%)	0	182 (9.02%)
Dropout	82 (4.07%)	196 (9.72%)	196 (9.72%)	167 (8.28%)	16 (0.79%)	657 (32.57%)
Current students	0	0	0	51 (2.53%)	82 (4.07%)	133 (6.59%)
Admission	397 (19.68%)	588 (29.15%)	532 (26.38%)	402 (19.93%)	98 (4.86%)	2,017 (100%)

Table 2 shows data compiled for the Business Computer Disciplines during the past twenty years (academic year 2001-2019). In which it is divided into five sets of data according to curriculum updates. In addition, the researcher presented a graph showing the consistency of the research problem in Figure 5.

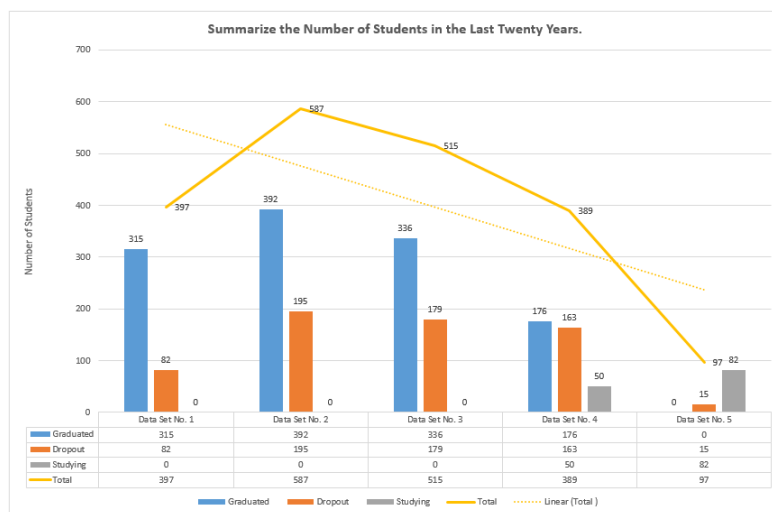


Fig. 5. Summarize the number of students in the last twenty years [16]

Figure 5 clearly shows that the overall student's tendency to decrease in each data set is shown in the dash line. While the solid line details the total number of students in each data set. Moreover, the bar graph showing the proportion of students who graduated and dropped out found that the trends of both parts were equal in volume (176:163 in the fourth series). In order to have more insight, the researchers explain in detail each set of data, as shown in Table 3 to Table 4.

Notation, the details of the non-graduate program (dropout) consist of the following three main points: Type 1 is problems leaving university due to timing and management, Type 2 is dropout with academic results, and Type 3 is retired.

Table 3. Students enrolled during academic year of 2001-2019

Year	Graduate		Dropout Classification			Total	Total Students
	Scheduled	Delayed	Type 1	Type 2	Type 3		
<i>Academic year of 2001-2013</i>							
2001	31 (79.49%)	1 (2.56%)	0	7 (100%)	0	7 (17.95%)	39 (9.82%)
2002	122 (84.14%)	8 (5.52%)	0	15 (100%)	0	15 (10.34%)	145 (36.52%)
2003	133 (62.44%)	20 (9.39%)	0	55 (91.67%)	5 (8.33%)	60 (28.17%)	213 (53.65%)
Total	286 (72.04%)	29 (7.30%)	0	77 (19.40%)	5 (1.26%)	82 (20.65%)	397 (100%)
<i>Academic year of 2004-2007</i>							
2004	116 (66.67%)	15 (8.62%)	2 (4.65%)	34 (79.07%)	7 (16.28%)	43 (24.71%)	174 (29.64%)
2005	83 (71.55%)	8 (6.90%)	1 (4.00%)	9 (36.00%)	15 (60.00%)	25 (21.55%)	116 (19.76%)
2006	73 (62.39%)	3 (2.56%)	4 (9.76%)	29 (70.73%)	8 (19.51%)	41 (35.59%)	117 (19.93%)

Year	Graduate		Dropout Classification				Total Students
	Scheduled	Delayed	Type 1	Type 2	Type 3	Total	
2007	84 (46.67%)	10 (5.56%)	11 (12.79%)	45 (52.33%)	30 (34.88%)	86 (47.78%)	180 (30.66%)
Total	356 (60.65%)	36 (6.13%)	18 (3.07%)	117 (19.93%)	60 (10.22%)	195 (33.22%)	587 (100%)
Academic year of 2008-2011							
2008	83 (48.54%)	29 (16.96%)	5 (8.47%)	41 (69.49%)	13 (22.03%)	59 (34.50%)	171 (33.20%)
2009	96 (67.61%)	13 (9.15%)	1 (3.03%)	20 (60.61%)	12 (36.36%)	33 (23.24%)	142 (27.57%)
2010	49 (47.57%)	8 (7.77%)	10 (21.74%)	24 (52.17%)	12 (26.09%)	46 (44.66%)	103 (20.00%)
2011	48 (48.48%)	10 (10.10%)	17 (41.46%)	8 (19.51%)	16 (39.02%)	41 (41.41%)	99 (19.22%)
Total	276 (53.59%)	60 (11.65%)	33 (6.41%)	93 (18.06%)	53 (10.29%)	179 (34.76%)	515 (100%)
Academic year of 2012-2016							
2012	38 (35.19%)	20 (18.52%)	9 (18.00%)	35 (70.00%)	6 (12.00%)	50 (46.30%)	108 (27.76%)
2013	29 (28.43%)	23 (22.55%)	15 (30.00%)	33 (66.00%)	2 (4.00%)	50 (49.02%)	102 (26.22%)
2014	26 (44.07%)	7 (11.86%)	3 (11.54%)	21 (80.77%)	2 (7.69%)	26 (44.07%)	59 (15.17%)
2015	33 (67.35%)	* Studying	8 (50.00%)	6 (37.50%)	2 (12.50%)	16 (32.65%)	49 (12.60%)
2016	* Studying	13 (61.90%)	6 (28.57%)	2 (9.52%)	21 (29.58%)	71 (18.25%)	
Total	126 (32.39%)	50 (13.44%)	48 (12.34%)	101 (25.96%)	14 (3.60%)	163 (41.90%)	389 (100%)
Academic year of 2017-2019							
2017	* Studying		1 (12.50%)	6 (75.00%)	1 (12.50%)	8 (24.24%)	33 (34.02%)
2018	* Studying		3 (42.86%)	0	4 (57.14%)	7 (20.59%)	34 (35.05%)
2019	* Studying		0	0	0	0	30 (30.93%)
Total	0	0	4 (26.67%)	6 (40.00%)	5 (33.33%)	15 (15.46%)	97 (100%)

From the data collected, it can be concluded that students of Business Computer Disciplines at the School of Information and Communication Technology, the University of Phayao have the most dropout during the 1st year (382 students or 60.25%). The reason that most students' dropout is due to their academic results (394 students or 19.85%). This is shown in Table 4.

Table 4. Details of students with dropout by class level during academic year of 2001-2019

Total Students Dropout: 634	Students' Dropped-out Year							
	1 st year	2 nd year	3 rd year	4 th year	5 th year	6 th year	7 th year	8 th year
	382 (60.25%)	190 (29.97%)	43 (6.78%)	11 (1.74%)	1 (0.16%)	5 (0.79%)	1 (0.16%)	1 (0.16%)
Total Students: 2,017	Graduate		Dropout Classification					
	Scheduled	Delayed	Type 1	Type 2	Type 3	Total		
	1,044 (52.59%)	175 (8.82%)	103 (5.19%)	394 (19.85%)	137 (6.90%)	634 (31.94%)		

From the data collected, it was found that the tendency for admission to the business computer disciplines tends to decrease significantly, as mentioned in Figure 6. The solid lines in Figure 6 show an increasing trend of student enrollment in 2003 and 2007, after 2008 the trend continued to decline clearly. In contract, the tendency of dropout shows the stability of the drop out as well as increases as shown in the dash line.

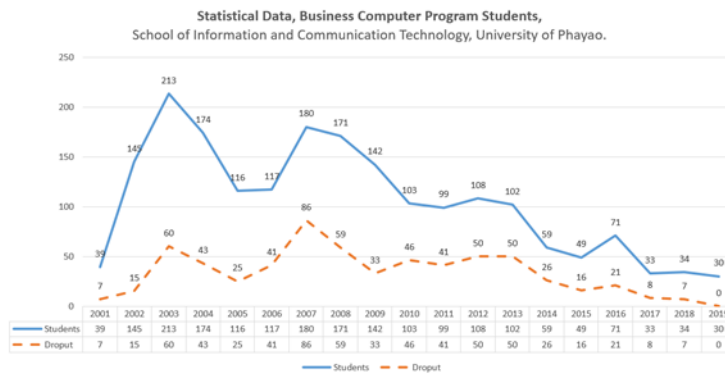


Fig. 6. Statistical data of business computer students during twenty years (2001-2019) [16]

It can therefore be concluded without any doubt that the situation should be resolved based on the students' academic performance results and finding preventive measures for first year students to stay in the program.

3.4 Modeling

From the data gathered, it is known that the number of students enrolled in the program tends to decrease continuously as displayed in Figure 6. In addition, the increasing trend of dropout reflects a number of problems related to the number of learners. It can be seen from the gap between the number of admissions and the number of dropouts which is necessary to find a solution to the problem urgently.

Therefore, this process presents a process for choosing the appropriate theory and machine learning tools to be used to develop the best solution model for suggestions

to the School of Information and Communication Technology, at the University of Phayao, Thailand.

The theory and machine learning tools chosen to develop the model were divided into two parts. The first part is the development of predictive models, it consisted of the Artificial Neural Network Algorithm (A-NN) [7], Decision Tree Algorithm (DT) [7][12], and Naïve Bayes Algorithm (NB) [7]. While the second part is the model performance testing as described in the evaluation section, which consisted of the confusion matrix performance, accuracy, precision, and recall measurement [8].

The model development process in this phase consisted of three steps. The first step is the researchers selecting and developing the model with machine learning techniques. The second step is that the researchers took the model from the first step to test the model's performance. Finally, the third step was to compare the performance of the model by selecting the appropriate model. An overview of modeling phase is shown in Figure 7.

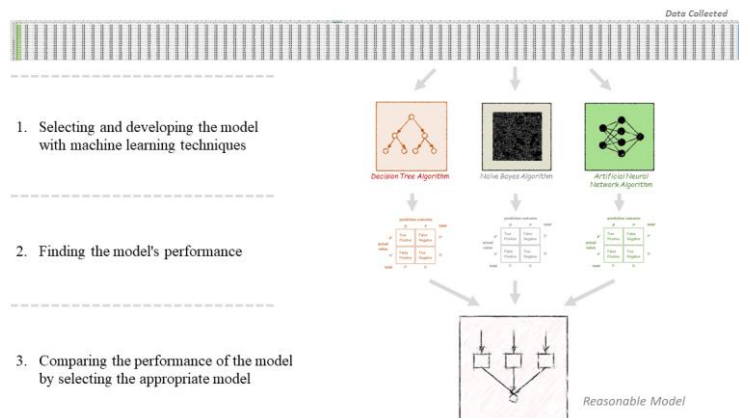


Fig. 7. An overview of modeling phase

3.5 Evaluation

This stage is part of the modeling phase which is the second step of model development. In principle, the performance evaluation of the model uses the cross-validation method [8]. The cross-validation method divides the data into two parts. The first part is used for model development and the second part is used for testing that model, as shown in Figure 8.

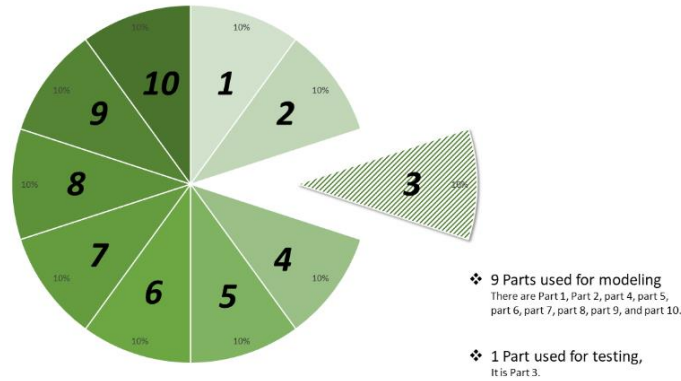


Fig. 8. Cross-Validation Methods [8]

Figure 8 describes the division of data for testing the model. In addition, model testing requires a tool called a confusion matrix [6], [17], [18] to test the model's performance. An important benefit of the performance of the confusion matrix is the ability to determine the model's ability to predict results, such as the predictive ability or accuracy, model precision, model sensitivity, and model specificity (recall measurement). These values are used to determine the actual performance model.

3.6 Deployment

Rüdiger Wirth said that in the CRISP-DM "Creation of the model is generally not the end of the project" [6]. Therefore, the researcher plans to deploy this research for being developed into a future recommendation application. An important goal is to test the real system with users in different universities.

4 Research Results

This research results and discussion phase has been divided into two stages as planned: Stage 1 is model performances and results, and Stage 2 is the best solution model and model selection.

4.1 Model performances and results

Models performance are the results of the testing by the confusion matrix performance, accuracy, precision, and recall measurement [8], which is separated by five research tools.

Modeling and results of the Artificial Neural Network Algorithm (A-NN): These results are part of a summary analysis for the selection of the reasonable mod-

els based on artificial neural network algorithms [7]. This process tests various patterns of cross-validation methods to find reasonable models with the highest accuracy.

Table 5. A-NN performance testing results

Academic Year	Cross-Validation Methods (CVM)			
	5-fold CVM	10-fold CVM	15-fold CVM	Leave-one-out CVM
2001-2003	89.04%*	87.47%	88.54%	86.73%
2004-2007	91.59%	92.32%	92.70%*	91.96%
2008-2011	93.71%	93.48%	93.44%	93.71%*
2012-2016	90.91%	90.94%	91.27%	91.64%*

Table 5 illustrates the four artificial neural network algorithm models that should be selected as follows: The suitable model for the student data enrolled in the academic year of 2001-2003 is the 5-Fold cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/36LU2d9>).

Table 6. Reasonable A-NN model’s performance during the academic year of 2001-2003

Predicted Condition	True Condition		Precision
	Graduated	Dropped	
Pred. Graduated	300	27	91.74%
Pred. Dropped	16	49	75.38%
Recall	94.94%	64.47%	

* Micro average: 89.03%

The suitable model for the student data enrolled in the academic year of 2004-2007 is the 15-Fold cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2R9xJHQ>).

Table 7. Reasonable A-NN model’s performance during the academic year of 2004-2007

Predicted Condition	True Condition		Precision
	Graduated	Dropped	
Pred. Graduated	381	27	93.38%
Pred. Dropped	13	126	90.65%
Recall	96.70%	82.35%	

* Micro average: 92.69%

The suitable model for the student data enrolled in the academic year of 2008-2011 is the leave-one-out cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2R8AB7L>).

Table 8. Reasonable A-NN model’s performance during the academic year of 2008-2011

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	324	14	95.86%
Pred. Dropped	13	78	85.71%
Recall	96.14%	84.78%	

* Micro average: 93.71%

The suitable model for the student data enrolled in the academic year of 2012-2016 is the leave-one-out cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2TeJXl3>).

Table 9. Reasonable A-NN model’s performance during the academic year of 2012-2016

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	163	11	93.68%
Pred. Dropped	12	89	88.12%
Recall	93.14%	89.99%	

* Micro average: 91.64%

Modeling and results of the Decision Tree Algorithm (DT): These results are part of a summary analysis for the selection of the reasonable models based on the Decision Tree Algorithm (DT) [15-16, 18]. The four models received have been summarized in Table 10 to Table 14, and the most effective model selection are concluded in Table 20 respectively.

Table 10. DT performance testing results

Academic Year	Model’s Depth	Cross-Validation Methods (CVM)			
		<i>5-fold CVM</i>	<i>10-fold CVM</i>	<i>15-fold CVM</i>	<i>Leave-one-out CVM</i>
2001-2003	Depth 2*	85.98%*	84.42%	85.18%	79.85%
	Depth 3	83.93%	84.92%	83.91%	81.89%
	Depth 4	83.68%	84.40%	84.17%	81.89%
	Depth 5	83.68%	84.40%	84.17%	81.89%
2004-2007	Depth 2	87.20%	87.20%	87.22%	87.20%
	Depth 3*	88.30%	87.74%	88.49%*	87.75%
	Depth 4	87.38%	87.01%	88.12%	87.57%
	Depth 5	87.38%	87.01%	88.12%	87.57%
2008-2011	Depth 2	82.76%	82.06%	80.85%	86.01%
	Depth 3	86.25%	85.33%	84.83%	89.74%
	Depth 4*	86.48%	85.33%	85.55%	89.98%*
	Depth 5	86.25%	85.33%	85.31%	89.98%
2012-2016	Depth 2*	81.45%	82.21%	84.05%	84.36%*
	Depth 3	83.27%	82.17%	82.51%	81.09%
	Depth 4	82.91%	82.91%	82.85%	82.18%
	Depth 5	82.91%	82.91%	82.85%	82.18%

Table 10 illustrates the four decision tree algorithm models that should be selected as follows: The suitable model for the student data enrolled in the academic year of 2001-2003 is the 5-Fold cross validation method, with the 2nd level depth of the decision tree model. In addition, the courses that have a significant influence on prediction in this model consist of one course: 001245 Science in Everyday Life. Table 11 shows the performance of the model.

Table 11. Reasonable DT model’s performance during the academic year of 2001-2003

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	297	36	89.19%
Pred. Dropped	19	40	67.80%
Recall	93.99%	52.63%	

* Micro average: 85.97 %

The suitable model for the student data enrolled in the academic year of 2004-2007 is the 15-Fold cross validation method, with the 3rd level depth of the decision tree model. In addition, the courses that have a significant influence on prediction in this model consist of four courses: 213100 Introduction to Business, 231110 Software Packages in Business, 231201 Business Statistics, and 231230 Data Structure and Algorithm. Table 12 shows the performance of the model.

Table 12. Reasonable DT model’s performance during the academic year of 2004-2007

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	372	41	90.07%
Pred. Dropped	22	112	83.58%
Recall	94.42%	73.20%	

* Micro average: 84.64%

The suitable model for the student data enrolled in the academic year of 2008-2011 is the leave-one-out cross validation method, with the 4th level depth of the decision tree model. In addition, the courses that have a significant influence on prediction in this model consist of six courses: 001103 Thai Language Skills, 001111 Foundations of English I, 001112 Foundations of English II, 001134 Civilization and Local Wisdom, 001173 Life Skills, and 231101 Business Statistics. Table 13 shows the performance of the model.

Table 13. Reasonable DT model’s performance during the academic year of 2008-2011

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	323	29	91.76%
Pred. Dropped	14	63	81.82%
Recall	95.85%	68.48%	

* Micro average: 89.98%

The suitable model for the student data enrolled in the academic year of 2012-2016 is the leave-one-out cross validation method, with the 2nd level depth of the decision tree model. In addition, the courses that have a significant influence on prediction in this model consist of one course: 221110 Fundamental Information Technology. Table 14 shows the performance of the model.

Table 14. Reasonable DT model’s performance during the academic year of 2012-2016

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	152	20	88.37%
Pred. Dropped	23	80	77.67%
Recall	86.86%	80.00%	

* Micro average: 84.73%

Modeling and results of the Naïve Bayes Algorithm (NB): These results are part of a summary analysis for the selection of the reasonable models based on the Naïve Bayes Algorithm (NB) [7]. This process tests various patterns of cross-validation methods to find reasonable models with the highest accuracy.

Table 15. NB performance testing results

Academic Year	Cross-Validation Methods (CVM)			
	<i>5-fold CVM</i>	<i>10-fold CVM</i>	<i>15-fold CVM</i>	<i>Leave-one-out CVM</i>
2001-2003	87.24%*	86.46%	85.95%	86.48%
2004-2007	90.31%	90.31%	90.67%*	90.31%
2008-2011	92.55%	93.24%	93.20%	93.47%*
2012-2016	90.91%	91.63%	91.68%*	91.64%

Table 15 illustrates the four Naïve Bayes algorithm models that should be selected as follows: The suitable model for the student data enrolled in the academic year of 2001-2003 is the 5-Fold cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2QHS08k>); they are included within the model in Table 14, and Table 15 shows the testing results of the model.

Table 16. Reasonable NB model’s performance during the academic year of 2001-2003

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	283	17	94.33%
Pred. Dropped	33	59	64.13%
Recall	89.56%	77.63%	

* Micro average: 87.24%

The suitable model for the student data enrolled in the academic year of 2004-2007 is the 15-Fold cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2sYSNJh>); they are included within the model in Table 44, and Table 41 shows the testing results of the model.

Table 17. Reasonable NB model’s performance during the academic year of 2004-2007

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	368	25	93.64%
Pred. Dropped	26	128	83.12%
Recall	93.40%	83.66%	

* Micro average: 90.68%

The suitable model for the student data enrolled in the academic year of 2008-2011 is the leave-one-out cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2QH2Mfd>); they are included within the model in Table 16, and Table 17 shows the testing results of the model.

Table 18. Reasonable NB model’s performance during the academic year of 2008-2011

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	323	14	95.85%
Pred. Dropped	14	78	84.78%
Recall	95.85%	84.78%	

* Micro average: 93.47%

The suitable model for the student data enrolled in the academic year of 2012-2016 is the 15-Fold cross validation method. The performance details and model prediction results are shown on the website (<https://bit.ly/2QKks9V>); they are included within the model in Table 18, and Table 19 shows the testing results of the model.

Table 19. Reasonable NB model’s performance during the academic year of 2012-2016

Predicted Condition	True Condition		Precision
	<i>Graduated</i>	<i>Dropped</i>	
Pred. Graduated	158	6	96.34%

Pred. Dropped	17	94	84.68%
Recall	90.29%	94.00%	

* Micro average: 91.64%

4.2 Best solution model and model selection

This section is a discussion of previous analysis results, with the objective of selecting the most reasonable model to present to relevant parties, School of Information and Communication Technology, and University of Phayao. Summary of analysis results in the previous section summarized in Table 20.

Table 20. Summary of analysis results

Academic Year	Modeling Analysis Results (Accuracy)		
	<i>A-NN</i>	<i>DT</i>	<i>NB</i>
2001-2003	89.04%*	85.98%	87.24%
2004-2007	92.70%*	88.49%	90.67%
2008-2011	93.71%*	89.98%	93.47%
2012-2016	91.64%	84.36%	91.68%*

5 Conclusion and Discussion

5.1 Research discussion

From Table 20, it can be concluded that the first model for the students enrolled during the academic year of 2001-2003 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 89.04% as shown model performance in Table 6. The second model for the students enrolled during the academic year of 2004-2007 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 92.70% as shown model performance in Table 7. The third model for the students enrolled during the academic year of 2008-2011 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 93.71% as shown model performance in Table 8. Finally, the last model for the students enrolled during the academic year of 2012-2016 is the Naïve Bayes Algorithm (NB) with an accuracy equal to 91.68% as shown model performance in Table 19.

5.2 Research conclusion

There are two main objectives of the research. The first objective is to construct reasonable students' dropout prediction model. The second objective is to evaluate the performance and select the best predictions of the dropout model. The data set used in the research is the 2,017-student data that has been enrolled for the past twenty years (2001-2019) from Business Computer Discipline at the School of Information and Communication Technology, at the University of Phayao. Research methodology is designed according to the data mining principle called CRISP-DM.

Research tools are divided into two main parts. The first part is the modelling section consisting of three types which are the Artificial Neural Network Algorithm (A-NN), Decision Tree Algorithm (DT), and Naïve Bayes Algorithm (NB). The second part is the model testing section, it consisted of the Confusion Matrix Performance, Accuracy, Precision, and Recall Measurement.

From the results, it can be concluded that the first model for the students enrolled during the academic year of 2001-2003 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 89.04%. The second model for the students enrolled during the academic year of 2004-2007 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 92.70%.

The third model for the students enrolled during the academic year of 2008-2011 is the Artificial Neural Network Algorithm (A-NN) with an accuracy equal to 93.71%. Finally, the last model for the students enrolled during the academic year of 2012-2016 is the Naïve Bayes Algorithm (NB) with an accuracy equal to 91.68%.

For the future, it is necessary to conduct additional research and present the research results to relevant parties and organizations for making critical decision and planning.

6 Acknowledgement

This research is supported by the three organizations: University of Phayao, Rajamangala University of Technology Tawan-Ok, and Rajabhat Mahasarakham University. The authors would like to thank the advisor, lecturers, students, technicians, and all respondents for their entire support.

7 References

- [1] M. Flavin, "Disruptive technologies in higher education," *Res. Learn. Technol.*, vol. 20, 2012.
- [2] S. Gallagher and G. Garrett, "Disruptive education: Technology-enabled universities," *U. S. Stud. Cent. Univ. Syd.*, 2013.
- [3] K. Phanniphong, P. Nuankaew, D. Teeraputon, W. Nuankaew, P. Tanasirathum, and S. Bussaman, "The Distinction Learning Style in Learning Outcomes of the Secondary School Learner," in 2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), Dec. 2018, pp. 1–5. <https://doi.org/10.1109/times-icon.2018.8621778>
- [4] K. Yadav, "Disruptive innovative technologies in higher education," *Int. J. Adv. Educ. Res.*, vol. 4, no. 1, pp. 49–54, Jan. 2019.
- [5] "Global Education Monitoring Report 2017/8: Accountability in education: Meeting our commitments - World," ReliefWeb. <https://reliefweb.int/report/world/global-education-monitoring-report-20178-accountability-education-meeting-our> (accessed Apr. 28, 2020). <https://doi.org/10.18356/95f6f839-en>
- [6] 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, Art. no. 21, Nov. 2019. <https://doi.org/10.3991/ijet.v14i21.11034>
- [7] M. Kang, H. Heo, I. Jo, J. Shin, and J. Seo, “Developing an educational performance indicator for new millennium learners,” *J. Res. Technol. Educ.*, vol. 43, no. 2, pp. 157–170, 2010.
- [8] 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.
- [9] 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>.
- [10] 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>
- [11] P. Nuankaew and P. Temdee, “Determining of compatible different attributes for online mentoring model,” in *2014 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronics Systems (VITAE)*, 2014, pp. 1–5. <https://doi.org/10.1109/vitae.2014.6934434>
- [12] P. Nuankaew, W. Nuankaew, K. Phanniphong, R. Fooprateepsiri, and S. Bussaman, “Analysis Dropout Situation of Business Computer Students at University of Phayao,” in *The Impact of the 4th Industrial Revolution on Engineering Education*, Cham, 2020, pp. 419–432. https://doi.org/10.1007/978-3-030-40274-7_42.
- [13] O. of the N. E. Commission, *Education in Thailand 2018*. Office of the National Education Commission Bangkok, Thailand, 2018.
- [14] P. Kantavong, T. Nethanomsak, and N. luang-ungkool, “Inclusive Education in Thailand after 1999 National Education Act: A Review of a Pre-Service Teacher Education System,” *Procedia - Soc. Behav. Sci.*, vol. 69, pp. 1043–1051, Dec. 2012. <https://doi.org/10.1016/j.sbspro.2012.12.031>.
- [15] P. Maneerat, K. Malaivongs, and J. Khlaisang, “The comparison of Thai qualification framework for higher education and capability maturity model integration for service,” *Procedia-Soc. Behav. Sci.*, vol. 182, pp. 225–231, 2015. <https://doi.org/10.1016/j.sbspro.2015.04.759>
- [16] P. Nuankaew, W. Nuankaew, D. Teeraputon, S. Bussaman and S. Rattarom, “Information Technology Educational Programs Situation in Disruptive Technology Era,” *Int. J. Adv. Sci. Technol.*, vol. 29, no. 05, pp. 4960–4984, May 2020.
- [17] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, “Confusion Matrix-based Feature Selection,” *MAICS*, vol. 710, pp. 120–127, 2011.
- [18] P. Nuankaew, W. Nuankaew, K. Phanniphong, S. Imwut, and S. Bussaman, “Students Model in Different Learning Styles of Academic Achievement at the University of Phayao, Thailand,” *Int. J. Emerg. Technol. Learn. IJET*, vol. 14, no. 12, pp. 133–157, Jun. 2019. <https://doi.org/10.3991/ijet.v14i12.10352>

8 Authors

Pratya Nuankaew received a B.Ed. Degree in Educational Technology in 2001, M.Sc. degree in Information Technology in 2008 from Naresuan University, and a Ph.D. degree in Computer Engineering in 2018 from Mae Fah Luang University. He is currently a lecturer at the School of Information and Communication Technology,

University of Phayao, Phayao, Thailand. His research interests are in online mentoring model, mentoring relationships, social network analysis, ubiquitous computing, learning analytics, digital education, and educational 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. Her research interests are in digital education, innovation and knowledge management, data science, and big data and information technology management.

Direk Teeraputon received his B.Ed. in Educational Technology in 1991 from the Silpakorn University, M.Ed. in Audio-Visual Education in 1995 and Ph.D. in Educational Technology and Communication in 2004 from the Chulalongkorn University. He is currently a Senior Lecturer at Faculty of Education, Naresuan University, Thailand. His research interests are instruction system design, learning resources center management, distance education, self-regulated learning and training.

Kanakarn Phanniphong received his B.B.A. degree in Information System in 2005 from Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand, M.B.A. degree Business Administration in 2008 and D.B.A. degree in Business in 2018 from Pathumthani University, Pathum Thani, Thailand. He is currently a lecturer at the Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Bangkok, Thailand. His research interests are, innovation and knowledge management, and management information system.

Sittichai Bussaman received his B.Sc. in Statistic in 1990 from Srinakharinwirot University, M.Sc. in Computer Science and Information Technology in 1997 from King Mongkut's Institute of Technology Ladkrabang and Ph.D. in Educational Technology and Communications in 2013 from Mahasarakham University. He is currently an Associate Professor at the Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand. His research interests are data mining in education, online learning, pattern recognition, and Artificial Intelligence.

Article submitted 2020-04-28. Resubmitted 2020-07-16. Final acceptance 2020-07-17. Final version published as submitted by the authors.