International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 15, Issue 3, 2021

Risk Management Models for Prediction of Dropout Students in Thailand Higher Education

Pratya Nuankaew[1*], Wongpanya Nuankaew[2], Patchara Nasa-ngium[3], [1]School of Information and Communication Technology, University of Phayao, Thailand, [2]Faculty of Information Technology, Rajabhat Mahasarakham University, Thailand, [3]Faculty of Science and Technology, Rajabhat Mahasarakham University, Thailand. Email address: pratya.nu@up.ac.th[1*], wongpanya.nu@up.ac.th[2], patchara.nu@rmu.ac.th[3]

This research aimed to study the insights of students' behavior over the past twenty years with three objectives. The first objective is to study the factors that influence the dropout of students in higher education. The second objective is to construct the model of the students' dropout prediction, and the third objective is to evaluate the performance of the risk management models. The collection includes the 2,042 instance of students who enrolled in a Business Computer Program at the School of Information and Communication Technology, University of Phayao. The research instruments were divided into three phases; Phase I is a basic analysis phase. It consists of number of students, number of courses, academic achievement results, and various summaries. Phase II is modeling phase. It consists of decision tree techniques, and feature selection methods. Phase III is model testing phase. It consists of cross-validation methods, confusion matrix performance, accuracy, precision, and recall measurements. The results of the research showed that all four models had the highest accuracy, including the 1st model with an accuracy of 88.15%, the 2nd model with an accuracy of 91.50%, the 3rd model with an accuracy of 87.97%, and the last model with an accuracy of 88.78%. In addition, the study found that there was a significant factor for all models by only one course. It is the Business Mathematics course which appears on all models. For future research, the key goal is to apply research findings to the curriculum design, provide appropriate teaching and learning, and manage sustainable tertiary solutions that will help solve learner dropout problems.

Key words: *Dropout Students, Digital Humanities, Educational Data Mining, Educational Engineering, Risk Management Model*

INTRODUCTION

In the age of technology which has had a wide influence on human life, the advancement of technology has resulted in constant and drastic changes in human life (Uzych, 2004). The readiness to create opportunities in various paths to support life consists entirely of technology to drive humanity. A prime example is the daily lifestyle, modern business, smart agriculture, artificial intelligence, as well as national economies and policies, all being intertwined and affected by changes in technology. These transformations lead people to change and adapt their lifestyle so that they can survive in an era of disruptive technology or finding the right balance from the impact (Lidicker, 2020).

Disruptive technology is modernization and innovation constructing new knowledge, new theories, new normal, develops new markets, and increases product value. It has a serious impact on the way people live and has replaced the original product or knowledge with changes in process, quality, efficiency, production, and sales (Abdel-Basset et al., 2021; Nagaraj, 2020). Moreover, the new knowledge-building process had been constantly modified, thus creating new products that differed from traditional manufacturing, resulting in drastic and even more radical changes. Examples of disruptive technology innovations include the development of artificial intelligence applied to forecasting, prediction, and hypothesis. In addition, automation is used in production processes, shipping processes, and storing data in cloud storage.

Disruptive technology will always influence human life, and inevitable with the transformation of wireless broadband and high-speed internet technology, big data analysis technology, robot technology, biotechnology, 3D printing technology, storage technology, energy storage and renewable energy technology. It can be seen that disruptive technology has set the precedent in the creation of a new outlook in career, knowledge, and on the way that the world should be viewed. Organizations can survive the changes in the technology age by studying and observing for continual changes in order to adapt to relevant technologies and respond to changes in the environment appropriately (Garrison, 2009). It is important for an organization to prepare and adapt itself to become better than they previously were.

In the dimensions of education (Combs & Meskó, 2015; Flavin, 2012), there are cooperation in various related sciences, including educational theories, educational technology, education and information technology, educational technology and communication, and computer education (P. Nuankaew & Temdee, 2019; W. Nuankaew & Nuankaew, 2019). The missing technology is the study and application of technologies related to Educational Data Mining (EDM) and Learning Analytics (LA) for generating new form of technologies, knowledge, and innovation. This approach is based on the concept of disruptive technology, which represents the New Educational Technology Domains Era Framework (NETDEF) expressing itself as the connection, relationship, and relevance of technology entering the new educational technology era, as shown in Figure 1.

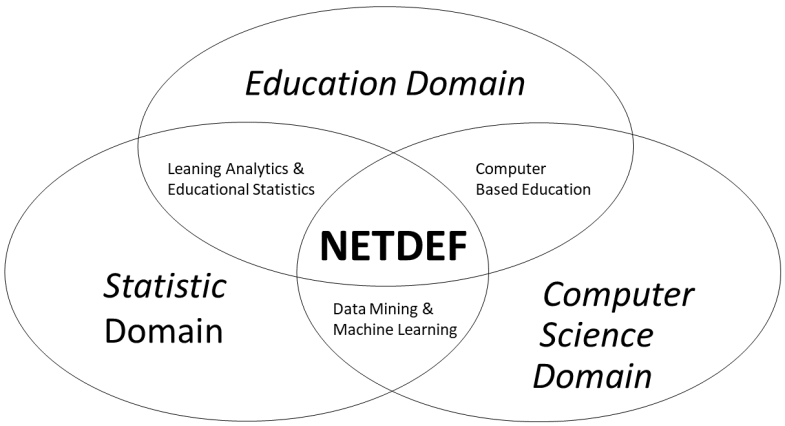


Figure 1. The New Educational Technology Domains Era Framework (NETDEF)

The New Educational Technology Domains Era Framework (NETDEF) lists the key ingredients of the framework: education domain, computer science domain, statistic domain, computer-based education domain, data mining and machine learning domain, and learning analytics and educational statistics domain. They are all elements and essential to the goal of establishing a research unit of excellence in Applied Informatics for Modernizations (Amis) at the School of Information and Communication Technology, University of Phayao.

Moreover, as a result of the drastically modified educational situation and context, it has a significant impact on all educational institutions in Thailand (Commission, 2006; Kantavong & Nethanomsak, 2012). Where the prime example, the major problem of the School of Information and Communications Technology at the University of Phayao, is in the number of dropout students increasing every academic year. Although the Business Computer program has been established and managed for over twenty years, it is not immune to experiencing a large number of undergraduate students dropping out of the program. Currently, students enrolled are less than a fifth from the past, however, the dropout rate was more than thirty percent in the first academic year. The case situation has therefore led to the determination of the research purpose which is presented in the next section.

Research Objectives

There are three significant objectives needed to solve student dropout problems in a Business Computer Program at the School of Information and Communication Technology, University of Phayao. The first objective is to study the factors that influence the dropout of students in higher education. The second objective is to construct the model of the students' dropout prediction, and the third objective is to evaluate the performance of the risk management models. The research dataset includes the 2,042 data set students enrolled in a Business Computer Program at the School of Information and Communication Technology, University of Phayao during the academic year 2001-2020.

Research Approach and Research Scope

The research approach was divided into three phases; The first phase is a basic analysis of the curriculum phase. It consists of a number of credits, number of courses, criteria for graduation, and various summaries. The second phase is the modelling phase. It consists of decision tree techniques and features the selection methods. Phase III is model testing phase. It consists of cross-validation methods, confusion matrix performance, accuracy, precision, and recall measurements.

The scope of the research is divided into five curriculum versions, with the curriculum being updated every 3-5 years according to national law. Five curriculum versions are comprised of the academic years of 2001-2020. The first curriculum was in the academic years of 2001-2003, the second curriculum was in the academic years of 2004-2007, the third curriculum was in the academic years of 2008-2011, the fourth curriculum was in the academic years of 2012-2016, and the last curriculum was in the academic years of 2017-2020 as detailed in Table 1.

The materials and methods of the research were followed by the Cross-Industry Standard Process for Data Mining Model, known as CRISP-DM (Chapman et al., 2000; Huber et al., 2019; Wirth & Hipp, 2000). It has a six-stage component: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Details of the implementation of the CRISP-DM model are presented in the next section.

METHODOLOGY

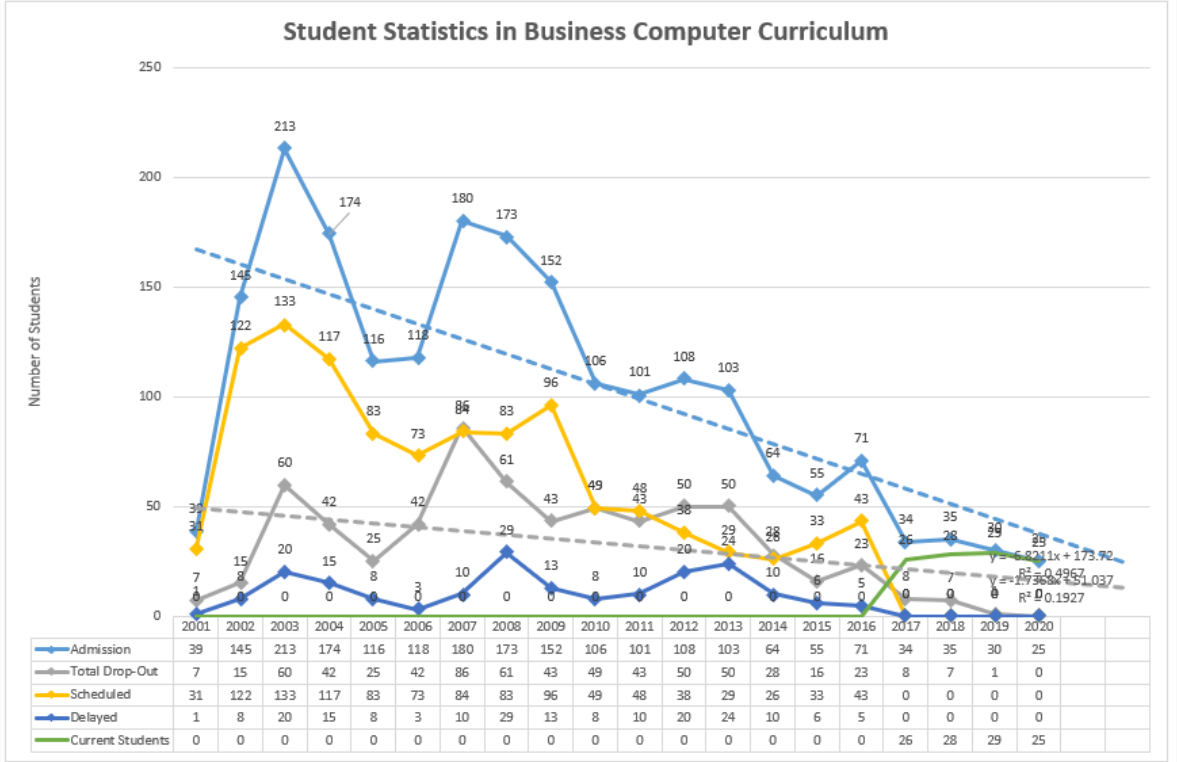
The elements and functions in this section were designed according to the principle of data mining using the CRISP-DM model. In addition, the CRISP-DM process is the process of creating knowledge from historical data and searching for insight into the data, known as Knowledge Discovery in Databases or KDD (Wirth & Hipp, 2000). The elements and functions in this section are designed according to data mining principles using the CRISP-DM model, where the six phases present the sequence of the content as follows:

Business Understanding

The key to understanding business problems is to explore the organization and find out what is the most important thing to solve first. In addition, the main principle is to discover the past, present and future impact of the organization (Chapman et al., 2000; Huber et al., 2019; Wirth & Hipp, 2000).

The problems of the research were to find out the root problems, the factors affecting the dropout of students, and the development of supporting models for managing the dropout problem of students in the Business Computer program at the School of Information and Communication Technology, University of Phayao. Initially, it was found that the problem of

students dropping out tended to increase while the number of learners tended to decrease, which is shown in Figure 2.



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Admission	39	145	213	174	116	118	180	173	152	106	101	108	103	64	55	71	34	35	30	25
Total Drop-Out	7	15	60	42	25	42	86	61	43	49	43	50	50	28	16	23	8	7	1	0
Scheduled	31	122	133	117	83	73	84	83	96	49	48	38	29	26	33	43	0	0	0	0
Delayed	1	8	20	15	8	3	10	29	13	8	10	20	24	10	6	5	0	0	0	0
Current Students	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	28	29	25

Figure 2. Data of Business Computer Students for twenty years (2001-2020)

The data collection illustrates the importance and potential future impact on the dropout of students in a Business Computer Program at the School of Information and Communication Technology, University of Phayao. Therefore, it is urgently necessary to conduct the research and study in order to plan a survival solution for the Business Computer Program.

Data Understanding

Understanding the data is relevant to understanding the research problems and preparing the data. The key points of understanding the data are to make plans for research data collection (Chapman et al., 2000; Wirth & Hipp, 2000). The four essential elements of understanding data are collecting the initial data, describing data, exploring data, and verifying the data.

In this paper, the researchers surveyed and gathered data to formulate the research problem presented in Figure 2. It was found that the current problem is the trend of program closing due to a lack of learners registering for the long term. For this reason, it is necessary to study the historical data in order to determine the root cause of the problem. The data collection is presented in Table 1 to Table 3.

Data Preparation

The purpose of this section is to collect good data and efficient (Chapman et al., 2000; Wirth & Hipp, 2000). Data collection is a research process through the steps and requirements of the University of Phayao. The process consists of obtaining research project approval, obtaining research ethical approval, obtaining data approval, and conducting a research. The data collected were 2,042 students in the Business Computer Program at the School of Information and Communication Technology, University of Phayao, Thailand. The data collections were divided into five datasets. The research data is made up of five data series as the program are updated every three to five years. The data for the research are classified and concluded in Tables 1 to Table 3.

Table 1. The five datasets from the data collection[1]

Datasets / Details	Graduation (Scheduled)	Graduation (Delayed)	Current Students	Dropped Out	Total
Academic year 2001 - 2003	286 (14.01%)	29 (1.42%)	0 (0.00%)	82 (4.02%)	397 (19.44%)
Academic year 2004 - 2007	357 (17.48%)	36 (1.76%)	0 (0.00%)	195 (9.55%)	588 (28.80%)
Academic year 2008 - 2011	276 (13.52%)	60 (2.94%)	0 (0.00%)	196 (9.60%)	532 (26.05%)
Academic year 2012 - 2016	169 (8.28%)	65 (3.18%)	0 (0.00%)	167 (8.18%)	401 (19.64%)
Academic year 2017 - 2020	0 (0.00%)	0 (0.00%)	108 (5.29%)	16 (0.78%)	124 (6.07%)
Total	1,088 (53.28%)	190 (9.30%)	108 (5.29%)	656 (32.13%)	2,042 (100.00%)

[1] Data updated: January 9, 2021.

Table 1 shows the details of the data collection. It comprises of five datasets: academic year 2001-2003 (397 students: 19.44%), academic year 2004-2007 (588 students: 28.80%), academic year 2008-2011 (532 students: 26.05%), academic year 2012-2016 (401 students: 19.64%), and academic year 2017-2020 (124 students: 6.07%). The data collected clearly show that the student dropout problem was very high: a proportion of 656 dropout students or equal to 32.13 percent. Details of the data collected for the academic year are shown in Table 2, and website: https://bit.ly/2XoPvKJ.

Table 2. The data collection

Curriculum / Version	Graduation (Scheduled)	Graduation (Delayed)	Current Students	Dropped Out	Total
Academic year 2001 – 2003					
2001	31	1	0	7	39
	(7.81%)	(0.25%)	(0.00%)	(1.76%)	(9.82%)
2002	122	8	0	15	145
	(30.73%)	(2.02%)	(0.00%)	(3.78%)	(36.52%)
2003	133	20	0	60	213
	(33.50%)	(5.04%)	(0.00%)	(15.11%)	(53.65%)
Total:	*286*	*29*	*0*	*82*	*397*
	(72.04%)	*(7.30%)*	*(0.00%)*	*(20.65%)*	*(100.00%)*
Academic year 2004 – 2007					
2004	117	15	0	42	174
	(19.90%)	(2.55%)	(0.00%)	(7.14%)	(29.59%)
2005	83	8	0	25	116
	(14.12%)	(1.36%)	(0.00%)	(4.25%)	(19.73%)
2006	73	3	0	42	118
	(12.41%)	(0.51%)	(0.00%)	(7.14%)	(20.07%)
2007	84	10	0	86	180
	(14.29%)	(1.70%)	(0.00%)	(14.63%)	(30.61%)
Total:	*357*	*36*	*0*	*195*	*588*
	(60.71%)	*(6.12%)*	*(0.00%)*	*(33.16%)*	*(100.00%)*
Academic year 2008 – 2011					
2008	83	29	0	61	173
	(15.60%)	(5.45%)	(0.00%)	11.47%)	(32.52%)
2009	96	13	0	43	152
	(18.05%)	(2.44%)	(0.00%)	8.08%)	(28.57%)
2010	49	8	0	49	106
	(9.21%)	(1.50%)	(0.00%)	(9.21%)	(19.92%)
2011	48	10	0	43	101
	(9.02%)	(1.88%)	(0.00%)	(8.08%)	(18.98%)
Total:	*276*	*60*	*0*	*196*	*532*
	(51.88%)	*(11.28%)*	*(0.00%)*	*(36.84%)*	*(100.00%)*
Academic year 2012 – 2016					
2012	38	20	0	50	108
	(9.48%)	(4.99%)	(0.00%)	(12.47%)	(26.93%)
2013	29	24	0	50	103
	(7.23%)	(5.99%)	(0.00%)	(12.47%)	(25.69%)
2014	26	10	0	28	64
	(6.48%)	(2.49%)	(0.00%)	(6.98%)	(15.96%)
2015	33	6	0	16	55
	(8.23%)	(1.50%)	(0.00%)	(3.99%)	(13.72%)
2016	43	5	0	23	71
	(10.72%)	(1.25%)	(0.00%)	(5.74%)	(17.71%)
Total:	*169*	*65*	*0*	*167*	*401*
	(42.14%)	*(16.21%)*	*(0.00%)*	*(41.65%)*	*(100.00%)*

Curriculum / Version	Graduation (Scheduled)	Graduation (Delayed)	Current Students	Dropped Out	Total
Academic year 2017 – 2020					
2017	0	0	26	8	34
	(0.00%)	(0.00%)	(20.97%)	(6.45%)	(27.42%)
2018	0	0	28	7	35
	(0.00%)	(0.00%)	(22.58%)	(5.65%)	(28.23%)
2019	0	0	29	1	30
	(0.00%)	(0.00%)	(23.39%)	(0.81%)	(24.19%)
2020	0	0	25	0	25
	(0.00%)	(0.00%)	(20.16%)	(0.00%)	(20.16%)
Total:	*0*	*0*	*108*	*16*	*124*
	(0.00%)	*(0.00%)*	*(87.10%)*	*(12.90%)*	*(100.00%)*

Table 2 details the data collected. It contains information on students who have completed their studies (graduate) and students who have dropped out of their studies, and students who have completed their studies. The data are classified into two sections: students who have completed their studies on schedule and those who have completed their studies delayed. The dropout students were classified into three categories, which consisted of type 1 dropout students with university conditions, type 2 dropout students with academic achievement criteria, and type 3 students who drop out by the reason for resigning from the institution. In addition, the data collected were also classified for the year level of the dropout students, which is shown in Table 3.

Table 3. Summarizing the data collection

Students Dropout: 656	**Students' Dropped-out Year**							
	1st year	**2nd year**	**3rd year**	**4th year**	**5th year**	**6th year**	**7th year**	**8th year**
	403	190	43	12	1	5	1	1
	(61.43%)	(28.96%)	(6.55%)	(1.83%)	(0.15%)	(0.76%)	(0.15%)	(0.15%)
Total Students: 2,042	**Current Students**	**Graduate**		**Dropout Classification**				
		Scheduled	**Delayed**	**Type 1**	**Type 2**	**Type 3**	**Total**	
	108	1,088	190	100	393	163	656	
	(5.29%)	(53.28%)	(9.30%)	(4.90%)	(19.25%)	(7.98%)	(32.13%)	

Table 3 shows that the major problem is that the number of dropout students is very high in Year 1. It had a large number of 403 students, representing 61.43 percent of the total number of dropout students. Moreover, the cause of dropout is also affected by the student's academic performance level. It has a total of 393 students, representing 19.25 percent of the total number of students.

In addition, data collection showed that the trend of admission to the Business Computer Program tends to decline significantly. It clearly countered the dropout of the student, which tended to rise as presented in Figure 2. The blue solid line in Figure 2 demonstrates an increasing trend in student enrolment during the academic year 2003-2007. After the academic year of 2008, the trend has decreased significantly. On the other hand, the trend of dropout shows the stability of dropout and an increase which is shown in the Gray solid line. 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 in an urgent manner.

Notation: Although the researcher collected all five sets of data, the fifth set during the 2017-2020 academic year was not yet complete due to the large number of students studying. For this reason, the researcher did not combine the aforementioned data for analysis in this research.

Modelling

Predictive modeling is a general concept in predictable modeling (Chapman et al., 2000; Wirth & Hipp, 2000). Such models typically have machine learning algorithms that learn certain properties from the training dataset to make those predictions. Predictive modeling can be divided into two subsections: regression and model classification. The regression model is based on the analysis of the relationship between variables and trends to make predictions about continuous variables. In contrast to the regression model, the task of classifying the model is to assign discrete class labels to a particular observation as the result of a prediction. For this reason, this research designed and selected the appropriate model development tools for the research, as shown in Figure 3.

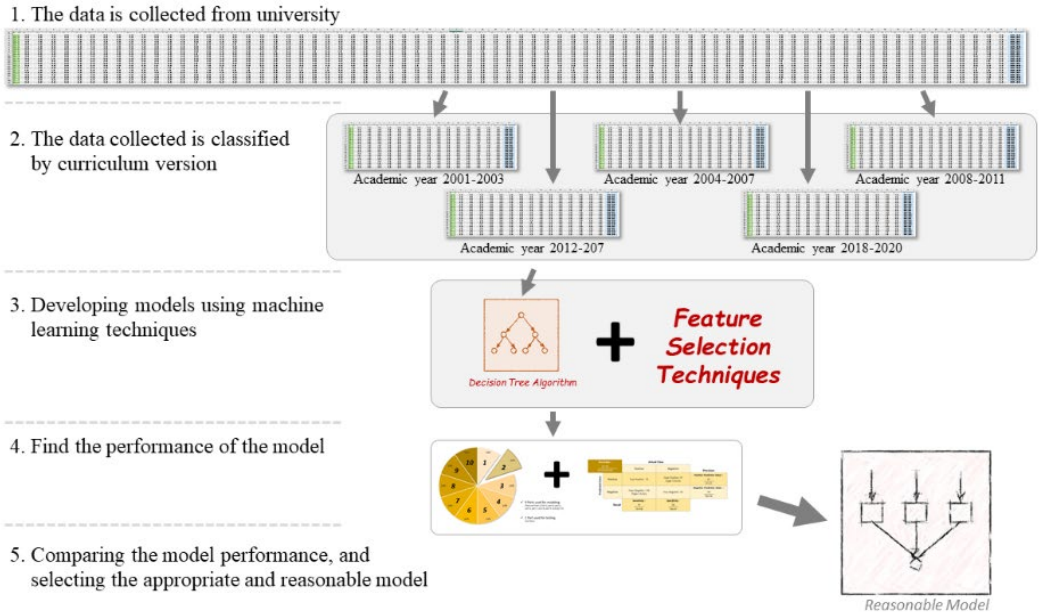


Figure 3. An Overview of Modeling Phase.

Figure 3 shows an overview of the development phase and model selection. It consists of 5 key processes: (1) Collection of data from the Business Computer Program at the University of Phayao, (2) Classification of data collected according to the curriculum versions. (3) Developing models using machine learning techniques. There are two types of tools used in the development of models. It consisted of Decision Tree classifiers (P. Nuankaew, 2020; W. Nuankaew & Nuankaew, 2020), and Feature Selection Techniques with Wrapper Method

(Chen & Chen, 2015; Rodriguez-Galiano et al., 2018). (4) Finding the performance of the model. (5) Comparing the model performance and selecting the reasonable model. The last two steps are described in the next section.

The principles of a decision tree classifier are very easy to understand. Imagine the decision tree is an upside-down tree. Where the top is the root and the lower part that cannot be separated anywhere is the leaf. It starts with a preliminary consideration of the starting point known as the root node if the data found meets that condition. The decision is then directed to the left of the root node to the point known as the child node, and if the data following this path meets the condition of this child node, it is terminated.

Feature selection is a technique that reduces the number of variables used in the development of prediction models. It can be performed to select a single optimal variable or a group of variables that are important for prediction. The feature selection process is an important process in the preparation of data mining. It is used to make prediction modeling more efficient as it reduces the dimensions and size of the data, which results in faster and higher quality learning methods.

Evaluation

Evaluation phase is part of the modeling phase which is the judgment step of model development (Chapman et al., 2000; Wirth & Hipp, 2000). In the process of preliminary judgment of the developed model, the model performance evaluation principle used the cross-validation method.

The cross-validation method is easy to analyze, as it divides the data into two parts. The first part is used for the constructed model, known as the training model. The second part is used to find out the performance of that model, known as the testing model. The conception of the cross-validation method is shown in Figure 4.

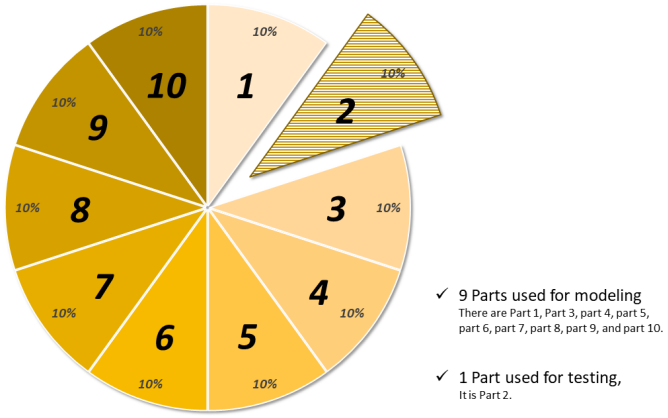


Figure 4. Cross-Validation Method

Figure 4 describes the division of data for testing the model, known as Cross-validation method. It separates the data into two parts: training model and testing model. This research has organized the data into two types of cross-validation methods. The 1st type of cross-validation method is 10-Fold cross-validation, which used 9-Fold (90 percent) for modeling and 1-Fold (10 percent) for testing. The 2nd type of cross-validation method is leave-one-out cross-validation, which used 99 percent of data for modeling and 1 percent of data for testing.

In addition, each time the cross-validation test is reported with the model results it was also tested by the confusion matrix performance. A great advantage on the performance of the confusion matrix is the ability to determine the model's ability to predict the results. The three major components of confusion matrix performance are measurement of accuracy, measurement of precision, and measurement of recall, which is shown as a formula in Figure 5.

Accuracy : $\frac{TP+TN}{TP+TN+FP+FN}$		**Actual Class**	
		Positive	*Negative*
Predicted Class	*Positive*	True Positive : TP	False Positive: FP (Type 1 Error)
	Negative	False Negative : FN (Type 2 Error)	True Negative: TN
Recall		***Sensitivity :*** $\frac{TP}{TP+FN}$	***Specificity :*** $\frac{TN}{TN+FP}$

Precision
Positive Predictive Value : $\frac{TP}{TP+FP}$
Negative Predictive Value : $\frac{TN}{TN+FN}$

Figure 5. Confusion Matrix Method

Deployment

Deployment is the implementation of a model, possibly building a system to achieve automation. In addition, deployment is also a research process to find knowledge or discoveries in what is being studied and researched (Chapman et al., 2000; Wirth & Hipp, 2000).

For solving a problem in this research, the researchers found that the number of enrolled students is declining sharply and the dropout rate among students is increasing at an alarming rate. The deployment should be a priority and emphasized on finding components, or risk creating a negative impact on students' academic achievement in the Business Computer Program at the School of Information and Communication Technology, University of Phayao should no corrective measures be taken at all.

RESULTS

The results of the research were divided into four sections according to the curriculum versions analyzed.

Academic year 2001-2003

The model results of the curriculum analysis in the academic year 2001-2003 have been shown in the model performance analysis in Table 4. While the selected model is shown in Table 5, and the significant attributes are shown in Table 6.

Table 4. The curriculum analysis in the academic year 2001-2003

Depth	Decision Tree		Decision Tree with Forward Selection		Decision Tree with Backward Elimination	
	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]
Depth 2	83.59%	85.89%*	85.41%	85.89%	86.64%	85.89%
Depth 3	83.87%	84.63%	87.94%	87.91%*	87.65%*	87.15%*
Depth 4	84.63%	83.88%	87.92%	87.91%	85.92%	86.15%
Depth 5	84.60%	84.13%	87.64%	87.91%	86.13%	86.15%
Depth 6*	85.13%*	84.13%	**88.15%***	87.91%	87.42%	86.15%
Depth 7	84.40%	84.13%	87.93%	87.91%	87.17%	86.15%
Depth 8	83.62%	84.13%	87.92%	87.91%	86.65%	86.15%
Depth 9	84.65%	84.13%	87.92%	87.91%	87.17%	86.15%

10-Fold[1] is 10-Fold cross-validation method. Leave-one-out[2] is Leave-one-out cross-validation method.

Table 4 shows the model performance analysis tests from the data set collected during the academic year of 2001-2003. It is divided into three methods: The first method is the model performance test with decision tree, the second method is the model performance test with decision tree and forward selection attributes, and the third method is the model performance test with decision tree and backward elimination attributes. For each test method, the researcher classified the cross-validation test into two parts: 10-Fold cross-validation and leave-one-out cross-validation. In addition, the researchers demonstrated the depth of the decision tree model in order to determine the suitable model to be deployed.

The results of the performance analysis of the model showed that the performance analysis with the second method has the highest accuracy. It is 88.15 % of accuracy with a depth 6 of the decision tree model and uses the 10-Fold cross-validation method, where the model performance analysis results are shown in Table 5. In addition, a summary of important characteristics is shown in Table 6.

Table 5. The performance of the selected model from the academic year 2001-2003

Accuracy: 88.15%		**Actual Class**		**Class Precision: 89.50%**
		True Graduated	**True Dropped**	
Predicted Class	**Pred. Graduated**	303	37	89.12%
	Pred. Dropped	12	45	78.95%
	Class Recall: 50.00%	96.19%	54.88%	

Table 5 shows the performance results of the selected model form the academic year of 2001-2003. The test results showed that it has an accuracy of 88.15 percent, class precision of 89.50 percent, and class recall of 50.00 percent.

Table 6. The significant attributes of the selected model from the academic year 2001-2003

No	Course ID	Course Name	Course Categories
1	001112	Foundations of English II	General Education Courses
2	001126	Thinking, Reasoning and Ethics	General Education Courses
3	001127	Man and Environment	General Education Courses
4	001135	Thai Studies	General Education Courses
5	001136	Global Issues	General Education Courses
6	231100	Business Mathematics	Specific Courses

Table 6 shows the significant attributes of the selected model from the academic year of 2001-2003. It consists of six courses from two categories of courses, with the vast majority of courses being most important are the general education courses.

Academic year 2004-2007

The model results of the curriculum analysis in the academic year of 2004-2007 have been shown in the model performance analysis in Table 7, while the selected model is shown in Table 8, and the significant attributes are shown in Table 9.

Table 7. The curriculum analysis in the academic year 2004-2007

Depth	Decision Tree		Decision Tree with Forward Selection		Decision Tree with Backward Elimination	
	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]
Depth 2	86.72%	86.73%	86.74%	86.73%	86.75%	86.73%
Depth 3	88.43%	89.29%	89.63%	89.80%	89.62%	89.63%
Depth 4*	88.61%	90.99%*	90.47%	90.31%*	90.65%	**91.50%***
Depth 5	89.63%	90.65%	90.80%	90.31%	90.98%	90.82%
Depth 6	88.93%	90.48%	90.97%	90.31%	90.64%	90.65%
Depth 7	90.31%*	90.48%	90.14%	90.31%	90.81%	90.65%
Depth 8	88.95%	90.48%	90.82%	90.31%	90.99%*	90.65%
Depth 9	89.12%	90.48%	90.98%*	90.31%	89.97%	90.65%

10-Fold[1] is 10-Fold cross-validation method. Leave-one-out[2] is Leave-one-out cross-validation method.

Table 7 shows the model performance analysis tests from the data set collected during the academic year of 2004-2007. It is divided into three methods: The first method is the model performance test with a decision tree, the second method is the model performance test with a decision tree and forward selection attributes, and the third method is the model performance test with a decision tree and backward elimination attributes. For each test method, the

researcher classified the cross-validation test into two parts: 10-Fold cross-validation and leave-one-out cross-validation. In addition, the researchers demonstrated the depth of the decision tree model in order to determine the suitable model to be deployed.

The results of the performance analysis of the model showed that the performance analysis with the third method was the highest accuracy. It is 91.50% of accuracy with a depth 4 of the decision tree model and uses the leave-one-out cross-validation method, where the model performance analysis results are shown in Table 8. In addition, a summary of important characteristics is shown in Table 9.

Table 8. The performance of the selected model from the academic year 2004-2007

Accuracy: 91.50%		**Actual Class**		**Class Precision: 90.35%**
		True Graduated	**True Dropped**	
Predicted Class	**Pred. Graduated**	154	9	94.48%
	Pred. Dropped	41	384	90.35%
	Class Recall: 97.71%	78.97%	97.71%	

Table 8 shows the performance results of the selected model form the academic year of 2004-2007. The test results showed that it has an accuracy of 91.50 percent, class precision of 90.35 percent, and class recall of 97.71 percent.

Table 9. The significant attributes of the selected model from the academic year 2004-2007

No	Course ID	Course Name	Course Categories
1	001103	Thai Language Skills	General Education Courses
2	001111	Foundations of English I	General Education Courses
3	001112	Foundations of English II	General Education Courses
4	001134	Conspectus of the Lower Northern Region	General Education Courses
5	001152	Body Conditioning	General Education Courses
6	001160	Human Behavior	General Education Courses
7	213100	Introduction to Business	Specific Courses
8	231100	Business Mathematics	Specific Courses
9	231110	Software Packages in Business	Specific Courses
10	231201	Business Statistics	Specific Courses
11	231230	Data Structure and Algorithm	Specific Courses

Table 9 shows the significant attributes of the selected model form the academic year of 2004-2007. It consists of eleven courses from two categories of courses, with the vast majority of courses being most important is the general education courses.

Academic year 2008-2011

The model results of the curriculum analysis in the academic year of 2008-2011 have been shown in the model performance analysis in Table 10, while the selected model is shown in Table 11, and the significant attributes are shown in Table 12.

Table 10. The curriculum analysis in the academic year 2008-2011

Depth	Decision Tree		Decision Tree with Forward Selection		Decision Tree with Backward Elimination	
	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]
Depth 2	83.11%	83.65%*	83.68%	83.65%	83.68%	83.65%
Depth 3	84.02%	81.95%	87.19%	86.65%	85.16%	85.34%
Depth 4	82.31%	83.08%	87.23%	87.59%	85.53%	87.22%*
Depth 5*	84.79%*	83.27%	86.65%	**87.97%***	86.09%	87.03%
Depth 6	84.40%	83.27%	87.23%	87.97%	86.66%	87.03%
Depth 7	83.08%	83.27%	87.22%	87.97%	86.85%	87.03%
Depth 8	84.79%	83.27%	87.22%	87.97%	86.84%	87.03%
Depth 9	83.28%	83.27%	87.40%*	87.97%	87.04%*	87.03%

10-Fold[1] is 10-Fold cross-validation method. Leave-one-out[2] is Leave-one-out cross-validation method.

Table 10 shows the model performance analysis tests from the data set collected during the academic year of 2008-2011. It is divided into three methods: The first method is the model performance test with a decision tree, the second method is the model performance test with a decision tree and forward selection attributes, and the third method is the model performance test with a decision tree and backward elimination attributes. For each test method, the researcher classified the cross-validation test into two parts: 10-Fold cross-validation and leave-one-out cross-validation. In addition, the researchers demonstrated the depth of the decision tree model in order to determine the suitable model to be deployed.

The results of the performance analysis of the model showed that the performance analysis with the second method has the highest accuracy. It is 87.97 of accuracy with a depth 5 of the decision tree model and uses the leave-one-out cross-validation method, where the model performance analysis results are shown in Table 11. In addition, a summary of important characteristics is shown in Table 12.

Table 11. The performance of the selected model from the academic year 2008-2011

Accuracy: 87.97%		**Actual Class**		**Class Precision: 91.77%**
		True Graduated	**True Dropped**	
Predicted Class	**Pred. Graduated**	323	51	86.36%
	Pred. Dropped	13	145	91.77%
	Class Recall: 73.98%	96.13%	73.98%	

Table 11 shows the performance results of the selected model form the academic year of 2001-2003. The test results showed that it has an accuracy of 87.97 percent, class precision of 91.77 percent, and class recall of 73.98 percent.

Table 12. The significant attributes of the selected model from the academic year 2008-2011

No	Course ID	Course Name	Course Categories
1	001111	Fundamental English	General Education Courses
2	001171	Life and Health	General Education Courses
3	213101	Introduction to Business	Specific Courses
4	231100	Business Mathematics	Specific Courses
5	213201	Business Management	Specific Courses
6	231130	Data Structure and Algorithm	Specific Courses

Table 12 shows the significant attributes of the selected model form the academic year of 2008-2011. It consists of six courses from two categories of courses, with the vast majority of the courses being most important is the specific courses.

Academic year 2012-2016

The model results of the curriculum analysis in the academic year of 2012-2016 have been shown in the model performance analysis in Table 13, while the selected model is shown in Table 14, and the significant attributes are shown in Table 15.

Table 13. The curriculum analysis in the academic year 2012-2016

Depth	Decision Tree		Decision Tree with Forward Selection		Decision Tree with Backward Elimination	
	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]	10-Fold[1]	Leave-one-out[2]
Depth 2	80.29%	78.55%	82.81%	82.79%	82.32%	82.79%
Depth 3	83.77%	85.54%	87.04%	86.78%	87.28%	86.78%
Depth 4	86.30%	85.04%	88.53%*	88.28%	87.52%	88.28%
Depth 5	87.04%	85.79%	88.03%	88.78%*	89.54%*	87.53%
Depth 6*	87.78%*	86.28%*	88.52%	88.78%	88.04%	**88.78%***
Depth 7	84.54%	86.28%	88.04%	88.78%	88.29%	88.78%
Depth 8	87.53%	86.28%	88.04%	88.78%	89.02%	88.78%
Depth 9	85.29%	86.28%	87.54%	88.78%	88.79%	88.78%

10-Fold[1] is 10-Fold cross-validation method. Leave-one-out[2] is Leave-one-out cross-validation method.

Table 13 shows the model performance analysis tests from the data set collected during the academic year of 2012-2016. It is divided into three methods: The first method is the model performance test with a decision tree, the second method is the model performance test with a decision tree and forward selection attributes, and the third method is the model performance test with a decision tree and backward elimination attributes. For each test method, the researcher classified the cross-validation test into two parts: 10-Fold cross-validation and

leave-one-out cross-validation. In addition, the researchers demonstrated the depth of the decision tree model in order to determine the suitable model to be deployed.

The results of the performance analysis of the model showed that the performance analysis with the third method has the highest accuracy. It is 88.78% of accuracy with a depth 6 of the decision tree model and uses the leave-one-out cross-validation method, where the model performance analysis results are shown in Table 14. In addition, a summary of important characteristics is shown in Table 15.

Table 14. The performance of the selected model from the academic year 2012-2016

Accuracy: 88.78%		**Actual Class**		**Class Precision: 86.21%**
		True Graduated	**True Dropped**	
Predicted Class	**Pred. Graduated**	131	9	93.57%
	Pred. Dropped	36	225	86.21%
	Class Recall: 96.15%	78.44%	96.15%	

Table 14 shows the performance results of the selected model form the academic year of 2012-2016. The test results showed that it has an accuracy of 88.78 percent, class precision of 86.21 percent, and class recall of 96.15 percent.

Table 15. The significant attributes of the selected model from the academic year 2012-2016

No	Course ID	Course Name	Course Categories
1	001103	Thai Language Skills	General Education Courses
2	003134	Civilization and Indigenous Wisdom	General Education Courses
3	004152	Body Conditioning	General Education Courses
4	005171	Life and Health	General Education Courses
5	005173	Life Skills	General Education Courses
6	122130	Principles of Management	Specific Courses
7	126100	Introduction to Economics	Specific Courses
8	128221	Principles of Marketing	Specific Courses
9	221100	Business Mathematics	Specific Courses
10	221110	Fundamental Information Technology	Specific Courses

Table 15 shows the significant attributes of the selected model form the academic year of 2012-2016. It consists of ten courses from two categories of courses, with the vast majority of the courses being most important are the general education courses, and specific courses.

DISCUSSION

In the discussion section of this research, the researchers conducted a debate by the three main objectives to solve the student dropout problems in the Business Computer Program at the School of Information and Communication Technology, University of Phayao. The first

objective is to study the factors that influence the dropout of students in higher education. The second objective is to construct the model of the students' dropout prediction, and the third objective is to evaluate the performance of the risk management models. Perspectives and discussion of the results for each objective are presented below.

Factors for Graduation and Dropout

The factors influencing graduation or dropout are the courses in which the learner has enrolled and achieved academic achievement each semester. In addition, the researchers found that the periods during which the students had the highest number of dropouts were academic year 1, as shown in Tables 1 to Tables 3, and website: https://bit.ly/2XoPvKJ.

The study found that the number of dropout students reached 656 students. Where the first academic year has 403 dropout students or equal to 61.43%, the second academic year has 190 dropout students or equal to 28.95%, the third academic year has 43 dropout students or equal to 6.55%, the fourth academic year has 12 dropout students or equal to 1.83%, the fifth academic year has 1 dropout students or equal to 0.15%, the sixth academic year has 5 dropout students or equal to 0.76%, the seventh academic year has 1 dropout students or equal to 0.15%, and the eighth academic year has 1 dropout students or equal to 0.15%.

According to the findings, most students dropped out in the first academic year due to a low level of academic achievement as shown in Table 3. It is especially relevant to the courses for learners enrolled in the 1st academic year. Therefore, it can be concluded that the courses enrolled in the 1st academic year have significant implications for students' academic achievement in the Business Computer Program at the School of Information and Communication Technology, University of Phayao. That course has been applied to develop models for predicting the students' academic achievement in the following section.

Model Construction

The compiled data that contains student data in the Business Computer Program consists of five data sets: academic year 2001-2003, academic year 2004-2007, academic year 2008-2011, academic year 2012-2006, and academic year 2017-2020, as shown the details on website: https://bit.ly/2XoPvKJ.

The researchers carried out an analysis to construct a model with the main goal of studying insight information of the past by classifying the analysis according to the updated curriculum. However, in the 2017-2020 academic year, the researcher did not take part in the analysis as there were still many students undergoing their studies. The analysis results are shown in Table 4 to Table 15, which can be summarized as the following:

The first curriculum had 397 students (19.44%) in the 2001-2003academic year, where the reasonable model was the Decision Tree model with forward selection attributes method. Its details include: the optimum depth of the decision tree is a depth 6, and the optimum cross-validation is 10-Fold cross-validation. The testing of the model performance result is equal to 88.15% of the accuracy value, 89.50% of the precision value, and 50.00% of the recall value as shown in Table 4, and Table 5. From such models, the researchers found that the courses were significant to the developed models. It consists of six courses including 001112 Foundations of English II, 001126 Thinking Reasoning and Ethics, 001127 Man and Environment, 001135 Thai Studies, 001136 Global Issues, and 231100 Business Mathematics as shown in Table 6.

The second curriculum had 588 students (28.80%) in the 2004-2007academic year, where the reasonable model was the Decision Tree model with backward elimination attributes method. Its details include: the optimum depth of the decision tree is a depth 4, the optimum cross-validation is leave-one-out cross-validation. The testing of the model performance result is equal to 91.50% of the accuracy value, 90.35% of the precision value, and 97.71% of the recall value as shown in Table 7, and Table 8. From such models, the researchers found that the courses were significant to the developed models. It consists of 11 courses including 001103 Thai Language Skills, 001111 Foundations of English I, 001112 Foundations of English II, 001134 Conspectus of the Lower Northern Region, 001152 Body Conditioning, 001160 Human Behavior, 213100 Introduction to Business, 231100 Business Mathematics, 231110 Software Packages in Business, 231201 Business Statistics, and 231230 Data Structure and Algorithm as shown in Table 9.

The third curriculum had 532 students (26.05%) in the 2008-2011academic year, where the reasonable model was the Decision Tree model with forward selection attributes method. Its details include: the optimum depth of the decision tree is a depth 5, the optimum cross-validation is leave-one-out cross-validation. The testing of the model performance result is equal to 87.97% of the accuracy value, 91.77% of the precision value, and 73.98% of the recall value as shown in Table 10, and Table 11. From such models, the researchers found that the courses were significant to the developed models. It consists of six courses including 001111 Fundamental English, 001171 Life and Health, 213101 Introduction to Business, 231100 Business Mathematics, 213201 Business Management, and 231130 Data Structure and Algorithm as shown in Table 12.

The fourth curriculum had 401 students (19.64%) in the academic year 2012-2016, where the reasonable model was the Decision Tree model with backward elimination attributes method. Its details include: the optimum depth of the decision tree is a depth 6, the optimum cross-validation is leave-one-out cross-validation. The testing of the model performance result is equal to 88.78% of the accuracy value, 86.21% of the precision value, and 96.15% of the recall value as shown in Table 13, and Table 14. From such models, the researchers found that the courses were significant to the developed models. It consists of 10 courses including 001103

Thai Language Skills, 003134 Civilization and Indigenous Wisdom, 004152 Body Conditioning, 005171 Life and Health, 005173 Life Skills, 122130 Principles of Management, 126100 Introduction to Economics, 128221 Principles of Marketing, 221100 Business Mathematics, and 221110 Fundamental Information Technology as shown in Table 15.

From the model construction and testing of the model, the results were significant courses to the prediction model, as summarized in Table 16.

Table 16. The summarized significant courses from the analysis models

Curriculum of Academic Year 2001-2003	Curriculum of Academic Year 2004-2007	Curriculum of Academic Year 2008-2011	Curriculum of Academic Year 2012-2016
001112	001103	001111	001103
001126	001111	001171	003134
001127	001112	213101	004152
001135	001134	**231100***	005171
001136	001152	213201	005173
231100*	001160	231130	122130
	213100		126100
	231100*		128221
	231110		**221100***
	231201		221110
	231230		
6 courses	11 courses	6 courses	10 courses

Table 16 shows a summary of the courses that are significant for each curriculum. The surprising finding of the research is that one course has significant implications for prediction in every course. It is a Business Mathematics with the following courses: 231100 Business Mathematics, and 221100 Business Mathematics. Even though the course code has changed, the course is the same course.

Evaluation of the Model Construction

The final part of the discussion is a summary of the model performance analysis, where the combination of the analysis results from the four selected models.

The data obtained from the four developed models, including the 1st model (academic year 2001-2003) had the model performance with an accuracy of 88.15%, precision of 89.50%, and recall of 50.00%. The 2nd model (academic year 2004-2007) had the model performance with an accuracy of 91.50%, precision of 90.35%, and recall of 97.71%. The 3rd model (academic year 2008-2011) had the model performance with an accuracy of 87.97%, precision of 91.77%, and recall of 73.98%. The 4th model (academic year 2021-2016) had the model performance with an accuracy of 88.78%, precision of 86.21%, and recall of 96.15%.

From the results, it was found that all models were highly effective. The findings in Table 16 revealed that one course is important to all models as it is part of the student's prediction of academic achievement. It is the Business Mathematics course, although the code in the fourth dataset has been changed, it is the same course in the course description. Thus, the conclusion from the research is that the Business Mathematics courses are important for planning future course developments in the Business Computer Program at the School of Information and Communication Technology, University of Phayao.

CONCLUSION

The dropout problem for tertiary students is a major problem affected by disruptive technology. Major disruptive technologies include social networking, mobile technology, network technology, and streaming entertainment, which have diminished learners' interest in the classroom, students' behaviour, and so on. Therefore, this research aimed to study the insights of students' behaviour from the past twenty years with three objectives. The first objective is to study the factors that influence the dropout of students in higher education. The second objective is to construct the model of the students' dropout prediction, and the third objective is to evaluate the performance of the risk management models.

The collection includes the 2,042 instance of students who enrolled in the Business Computer Program at the School of Information and Communication Technology, University of Phayao. It comprises of five datasets: The first set of data was 397 students (19.44%) who were enrolled in the 2001-2003academic year. The second set of data was 588 students (28.80%) who were enrolled in the 2004-2007academic year. The third set of data was 532 students (26.05%) who were enrolled in the 2008-2011academic year. The fourth set of data was 401 students (19.64%) who were enrolled in the 2012-2016 academic year. The last set of data was 124 students (6.07%) who were enrolled in the 2017-2020 academic year. The details of the data collected are shown and summarized in Table 1 to Table 4, and website: https://bit.ly/2XoPvKJ. The data collected clearly showed that the student dropout problem was very high, with the proportion of 656 dropping out or equal to 32.13 percent, as concluded in Table 1. Moreover, the researchers found that most of the students had the highest dropout problem in Year 1 as summarized in Table 3 and presented in the information on the website.

From the findings of data collection, the research instruments were divided into three phases; Phase I is a basic analysis phase. It consists of a number of students, number of courses, academic achievement results, and various summaries. Phase II is the modelling phase. It consists of decision tree techniques, and feature selection methods. Phase III is the model testing phase. It consists of cross-validation methods, confusion matrix performance, accuracy, precision, and recall measurements. The entire process is carried out according to the CRISP-DM principle, which consists of six key phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

The results of the research showed that all four models had the highest accuracy, including the 1st model with an accuracy of 88.15%, the 2nd model with an accuracy of 91.50%, the 3rd model with an accuracy of 87.97%, and the last model with an accuracy of 88.78%. In addition, the work found that there was a significant factor for all models by only one course. It is the Business Mathematics course which appeared on all models. For future research, the key goal is to apply research findings to the curriculum design, provide appropriate teaching and learning, and manage sustainable tertiary solutions that will help prevent undergraduates from dropping out of the program.

ACKNOWLEDGEMENT

This research is supported by two universities. It consists of the University of Phayao, and the Rajabhat Mahasarakham University. The authors would like to thank the advisor, lecturers, students, technicians, and all respondents for their entire support. This research was funded by the project of the Unit of Excellence in Applied Informatics for Modernization (FF64-UoE004) from the University of Phayao.

CONFLICT OF INTEREST AND RESEARCH ETHICS

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

REFERENCES

Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2021). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change*, *163*, 120431. https://doi.org/10.1016/j.techfore.2020.120431

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Inc*, *9*, 13.

Chen, G., & Chen, J. (2015). A novel wrapper method for feature selection and its applications. *Neurocomputing*, *159*, 219–226.

Combs, C. D., & Meskó, B. (2015). Disruptive technologies affecting education and their implications for curricular redesign. In *The Transformation of Academic Health Centers* (pp. 57–68). Elsevier.

Commission, O. of the H. E. (2006). National qualifications framework for higher education in Thailand implementation handbook. *Retrieved March*, *2*, 2014.

Flavin, M. (2012). Disruptive technologies in higher education. *Research in Learning Technology*, *20*. https://doi.org/10.3402/rlt.v20i0.19184

Garrison, G. (2009). An assessment of organizational size and sense and response capability on the early adoption of disruptive technology. *Computers in Human Behavior*, *25*(2), 444–449. https://doi.org/10.1016/j.chb.2008.10.007

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. *Procedia Cirp*, *79*, 403–408.

Kantavong, P., & Nethanomsak, T. (2012). Inclusive education in Thailand after 1999 National Education Act: A review of a pre-service teacher education system. *Procedia-Social and Behavioral Sciences*, *69*, 1043–1051.

Lidicker, W. Z. (2020). A Scientist's Warning to humanity on human population growth. *Global Ecology and Conservation*, *24*, e01232. https://doi.org/10.1016/j.gecco.2020.e01232

Nagaraj, S. V. (2020). Disruptive technologies that are likely to shape future jobs. *Procedia Computer Science*, *172*, 502–504.

Nuankaew, P. (2020). Clustering of Mindset towards Self-Regulated Learning of Undergraduate Students at the University of Phayao. *Advances in Science, Technology and Engineering Systems*, *5*(4), 676–685. https://doi.org/10.25046/aj050481

Nuankaew, P., & Temdee, P. (2019). Matching of compatible different attributes for compatibility of members and groups. *International Journal of Mobile Learning and Organisation*, *13*(1), 4–29. https://doi.org/doi.org/10.1504/IJMLO.2019.096469

Nuankaew, W., & Nuankaew, P. (2019). The Study of the Factors and Development of Educational Model: The Relationship between the Learner Context and the Curriculum Context in Higher Education. *International Journal of Emerging Technologies in Learning (IJET)*, *14*(21), 205–226. https://doi.org/10.3991/ijet.v14i21.11034

Nuankaew, W., & Nuankaew, P. (2020). Tolerance of Characteristics and Attributes in Developing Student's Academic Achievements. *Advances in Science, Technology and Engineering Systems Journal*, *5*(5), 1126–1136. https://doi.org/10.25046/aj0505137

Rodriguez-Galiano, V. F., Luque-Espinar, J. A., Chica-Olmo, M., & Mendes, M. P. (2018). Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Science of The Total Environment*, *624*, 661–672. https://doi.org/10.1016/j.scitotenv.2017.12.152

Uzych, L. (2004). Between technology and humanity: The impact of technology on healthcare ethics: Chris Gastmans (Ed.), Leuven University Press, Leuven (Belgium), 2002, 259pp, 37 Euro (paper). *Social Science & Medicine*, *58*(4), 879. https://doi.org/10.1016/S0277-9536(03)00235-1

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