Improving Predictive Model to Prevent Students' Dropout in Higher Education Using Majority Voting and Data Mining Techniques

Pratya Nuankaew[1] , Patchara Nasa-Ngium[2] , and Wongpanya S. Nuankaew[2(✉)]

[1] University of Phayao, Phayao 56000, Thailand
[2] Rajabhat Maha Sarakham University, Maha Sarakham 44000, Thailand
wongpanya.nu@rmu.ac.th

Abstract. The primary objective of this research was to improve the predictive model to prevent dropouts among university students. There were two secondary research objectives: (1) to study the context and improve the student dropout prevention model and (2) to compare the past university student dropout models. The research population was students in the Business Computer Department at the School of Information and Communication Technology, University of Phayao. A research tool was a model development process using majority voting and data mining techniques. The results showed that the model for predicting dropout prevention among university students was more effective. The model obtained was 83.62% accurate with 3-ensemble majority voting, including Generalized Linear Model (GLMs), Neural Network (NN), and Decision Tree (DT). The F1-Score for the dropped and scheduled graduation class was very high with 99.57% and 81.82%. The model derived from this research improved efficiency and predicted student dropout at the university level better than the previous model. Therefore, in future curriculum improvements, method matter issues that influence the dropout of university students should be considered.

Keywords: Dropout preventing · Educational data mining · Majority voting · Students dropout model · University dropout

1 Introduction

A university is an educational institution that allows students to learn to develop themselves into opportunities and career possibilities in the future. The university is responsible for producing quality graduates with the skills and potential to apply the knowledge gained and develop them appropriately for their prospective careers. In any case, the potential and learning styles of students at universities affect students' learning achievement. Numerous studies have compiled a list of phenomena that affect tertiary students who unexpectedly drop out of the education system by various factors: academic exhaustion, satisfaction with education, willingness to dropout, academic achievement performance, funding, and disabilities [1–4]. A fundamental problem discovered by many studies is that most university students have dropout problems in their first year

of study [1, 2, 5]. Additionally, dropout issues were highlighted, including data scientists and artificial intelligence. The use of educational data mining and machine learning to encounter solutions for student dropouts has become a research area that combines scientific and social science knowledge [2, 6–8]. The dropout problem among students at all educational levels is an academic and social waste that educators and scientists should not pass on.

Researchers in this research, who are responsible for curricula management at the University of Phayao, face the problem of dropouts among many students. Therefore, it is necessary to solve the problem of dropouts among university students. In the past, Pratya Nuankaew [9] has developed models using decision tree classification techniques. He found that the developed model had an accuracy of 87.21%, predicting two aspects of learning achievement. That research also showed weaknesses with the development of only one modeling technique. Subsequently, Nuankaew et al. [10] jointly develop further research to select techniques for developing models that can predict student achievement in more diverse programs. They have added three classification techniques to compare and select the best model with the highest accuracy. They found the Naïve Bayes technique to be the most accurate, with an accuracy of 91.68%, which the Department of Business Computer can use to plan the prevention of dropouts among current students effectively. However, the dropout problem has now been eliminated, but the problem of delayed graduation is increasing among the next generation of students. Therefore, this research aimed to create a predictive model for preventing dropouts and predicting a group of students with a chance of delaying graduation. There are two objectives of the research. The first objective is to study the context and improve the student dropout prevention model at the university level. The second objective is to compare the past university students' dropout models. The research population was students in the Department of Business Computing, School of Information and Communication Technology, the University of Phayao during the academic year 2012–2016. The research tool offers a new and sophisticated approach called majority voting. In addition, the researchers opted for a more diverse prediction technique, including Generalized Linear Model (GLMs), Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), and k-Nearest Neighbor (k-NN). Finally, the researchers used cross-validation techniques and confusion matrix assessment to assess the model's effectiveness.

Therefore, the researcher is interested in studying to prevent students from dropping out and delaying graduation to formulate a strategic plan for the next generation of educational administration. Researchers firmly believe that this research will significantly impact on improving the quality of education.

2   Materials and Methods

2.1   Population and Sample

The research population was students in the Department of Business Computing, School of Information and Communication Technology, the University of Phayao during the academic year 2012–2016. The data used as a research sample were students who had registered and received academic results in the Bachelor of Business Administration program in Business Computer.

Research samples are summarized and classified by academic year as presented in Table 1.

Table 1. Data collection

Academic year	Total students	Graduated		Dropped out
		Scheduled	Delayed	
2012–2015	99 (28.45%)	32 (9.20%)	26 (7.47%)	41 (11.78%)
2013–2016	92 (26.44%)	28 (8.05%)	24 (6.90%)	40 (11.49%)
2014–2017	57 (16.38%)	24 (6.90%)	12 (3.45%)	21 (6.03%)
2015–2018	46 (13.22%)	33 (9.48%)	6 (1.72%)	7 (2.01%)
2016–2019	54 (15.52%)	31 (8.91%)	17 (4.89%)	6 (1.72%)
Total	348 (100%)	148 (42.53%)	85 (24.43%)	115 (33.05%)

Table 1 presents a summary of data collection for research purposes. It contains 5 data sets of students in the Business Computer Program from the School of Information and Communication Technology at the University of Phayao. The data in Table 1 showed that the overall number of students decreased. There are also three points of interest: the number of graduates as scheduled is only 42.53%, while the number of graduates as delayed is 24.43%. There are as many as 33.05% of students who drop out. Therefore, research is a reason for developing predictive models to prevent students dropouts.

2.2 Data Acquisition Procedure

The data acquisition process consists of five phases. The first phase is a process of studying the feasibility and problems of research. Researchers found that students in the Business Computer Program from the School of Information and Communication Technology at the University of Phayao continued to decline. It also covers the issue of delayed graduation and dropout as the main problem. Researchers were given the policy in the second phase to find solutions that led to the research problem. The research problem is what factors affect the student's academic achievement? In Phase 3, researchers have requested human research ethics, which the University of Phayao has approved (UP-HEC: 2/020/63). In Phase 4, researchers coordinated to request academic achievement data from the University of Phayao, which received 254,456 transactions of student achievement data.

The researchers kept the information confidential and not disclosed according to the regulations of the University of Phayao. In Phase 5, researchers extracted data to prepare an analysis for model development. Researchers classified the data into three groups. The first group was students who graduated as scheduled, the second group was students who graduated as delayed, and the last group was dropped out students.

Furthermore, the researchers found that many students dropped out in the first academic year, as shown in Table 2.

Table 2. Student dropout statistics

Academic year	Total dropped out	Classified by academic year level				
		1st year	2nd year	3rd year	4th year	More 4
2012–2015	41 (35.65%)	23 (20.00%)	12 (10.43%)	3 (2.61%)	1 (0.87%)	2 (1.74%)
2013–2016	40 (34.78%)	21 (18.26%)	10 (8.70%)	5 (4.35%)	1 (0.87%)	3 (2.61%)
2014–2017	21 (18.26%)	13 (11.30%)	7 (6.09%)	1 (0.87%)	0	0
2015–2018	7 (6.09%)	5 (4.35%)	2 (1.74%)	0	0	0
2016–2019	6 (5.22%)	5 (4.35%)	1 (0.87%)	0	0	0
Total	115 (100%)	67 (58.26%)	32 (27.83%)	9 (7.83%)	2 (1.74%)	5 (4.35%)

Table 2 clearly shows that the dropout problem is significant. The students enrolled in 1st year had the highest number of dropouts, with 67 students (58.26%). Students in the 2nd year have the second dropout number, with 32 students (27.83%). For this reason, the researchers limited the scope of the first-year academic achievement data to create a predictive model to prevent students' dropout in the Bachelor of Business Administration program in Business Computer at the School of Information and Communication Technology, the University of Phayao.

2.3 Model Construction Tools

This section aims to design machine learning tools to construct predictive models to prevent students' dropout in the Bachelor of Business Administration program in Business Computer at the School of Information and Communication Technology, the University of Phayao. In the past, Pratya Nuankaew [9] has developed models using decision tree classification techniques. He found that the developed model had an accuracy of 87.21%, predicting two aspects of learning achievement. That research also showed weaknesses with the development of only one modeling technique. Subsequently, Nuankaew et al. [10] jointly developed further research to select techniques for developing models that can predict student achievement in more diverse programs. They have added three classification techniques to compare and select the best model with the highest accuracy. They found the Naïve Bayes technique to be the most accurate, with an accuracy of 91.68%, which the Department of Business Computer can use to plan the prevention of dropouts among current students effectively. However, the dropout problem has now been eliminated, but the problem of delayed graduation is increasing among the next generation of students.

Therefore, this research aimed to create a predictive model for preventing dropouts and predicting a group of students with a chance of delaying graduation. As mentioned above, the predictive model class consists of three domains: scheduled graduation, delayed graduation, and dropout. The researchers used the majority voting technique to select the most efficient models for more outstanding performance. The model development framework is presented in Fig. 1.
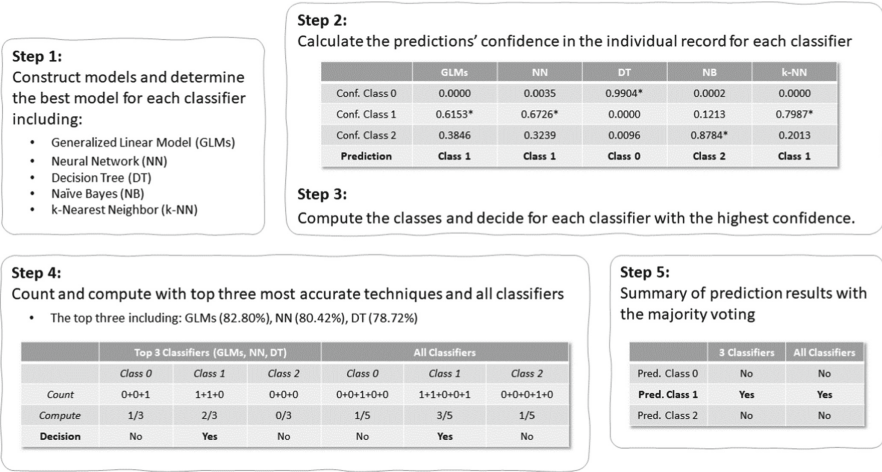
Step 1:
Construct models and determine the best model for each classifier including:
- Generalized Linear Model (GLMs)
- Neural Network (NN)
- Decision Tree (DT)
- Naïve Bayes (NB)
- k-Nearest Neighbor (k-NN)

Step 2:
Calculate the predictions' confidence in the individual record for each classifier

	GLMs	NN	DT	NB	k-NN
Conf. Class 0	0.0000	0.0035	0.9904*	0.0002	0.0000
Conf. Class 1	0.6153*	0.6726*	0.0000	0.1213	0.7987*
Conf. Class 2	0.3846	0.3239	0.0096	0.8784*	0.2013
Prediction	Class 1	Class 1	Class 0	Class 2	Class 1

Step 3:
Compute the classes and decide for each classifier with the highest confidence.

Step 4:
Count and compute with top three most accurate techniques and all classifiers
- The top three including: GLMs (82.80%), NN (80.42%), DT (78.72%)

	Top 3 Classifiers (GLMs, NN, DT)			All Classifiers		
	Class 0	Class 1	Class 2	Class 0	Class 1	Class 2
Count	0+0+1	1+1+0	0+0+0	0+0+1+0+0	1+1+0+0+1	0+0+0+1+0
Compute	1/3	2/3	0/3	1/5	3/5	1/5
Decision	No	Yes	No	No	Yes	No

Step 5:
Summary of prediction results with the majority voting

	3 Classifiers	All Classifiers
Pred. Class 0	No	No
Pred. Class 1	Yes	Yes
Pred. Class 2	No	No

Fig. 1. The model development framework

There are five steps to improving predictive models to prevent student dropouts in higher education using majority voting and data mining techniques. The first step is constructing models and determining the best model for each classifier. The selected classifiers consisted of five techniques: Generalized Linear Model (GLMs), Neural Network (NN), Decision Tree (DT), Naïve Bayes (NB), and k-Nearest Neighbor (k-NN). The models for each classifier chosen are present in Table 3. The second step is calculating the prediction confidence in each record for the classifier. The calculations in step 2 aim to find the conclusions for each prediction of each technique. The third step is to decide the answer based on the highest confidence value.

The fourth step is the crucial step of the process. This step is divided into two parts and four sub-steps. The first part was to consider the majority vote with the top three most accurate models. The second part uses all the models developed to determine the vote. The four sub-steps of both sections perform the same: counting the vote statistics, calculating the stats divided by the number of classifiers, averaging, and deciding a reasonable class. Step 4 in Fig. 1 describes this process. The fifth step summarizes the majority vote and compares the decision of the two parts in Step 4. As shown in Step 5, an example of a comparison of the two parts is shown, indicating that the votes of both parts give the same class.

2.4 Model Performance Evaluation Tools

The purpose of model performance evaluation is to verify the validity obtained from the model's predictive results compared to the actual data. The techniques decide to assess the effectiveness of the model in this work. It consists of two approaches: the cross-validation technique and the confusion matrix assessment [11].

Principles and testing of cross-validation technique consist of dividing the data into two parts. The first part used to create the model is called the training dataset. The rest used to test the model is called the testing dataset. The workflow of the cross-validation

technique consists of five steps: The first step is to divide the random data set into training and testing datasets. The second step is to put the model on the training dataset. The third step is to test the model with the testing dataset. The fourth step is calculating the accuracy of statistics using the testing dataset (Step 3). The final step repeats steps 1 to 4 and averages the results.

A confusion matrix is a method used for evaluating the performance of a classification model where the number of target classes is the dimension of the upcoming matrix. The tool used as a model's performance index, coupled with a confusion matrix, consists of four indicators. The first indicator is accuracy, calculated by the number of correctly predicted results divided by the total amount of data. The second indicator is precision, which tells us how many cases are accurately predicted in the class of interest. The third indicator is recall, which tells us how many cases are accurately predicted in the actual class. The last indicator is F1-Score, which shows performance by taking the precision and recall values to calculate the mean, called Harmonic Mean. The composition and calculation of each indicator is shown in Fig. 2.
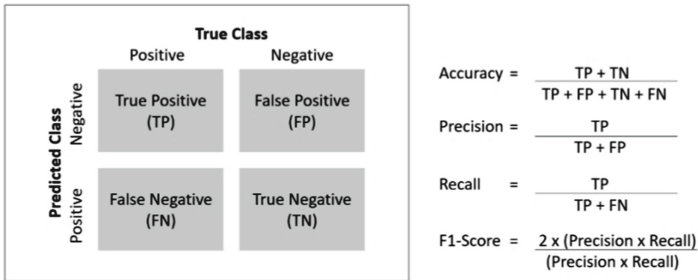

Fig. 2. The elements and calculations in the confusion matrix

This research used cross-validation and confusion matrix techniques in two phases. The first phase evaluates model performance for each classifier in step 2 of the research framework, and the second phase estimates the model's performance using the majority voting technique in step 5. Dividing the data for testing and evaluating in Step 2 and Step 5 consisted of two types of cross-validation: 10-Fold and Leave-one-out cross-validation. The best results of testing and evaluating model performance are shown in Tables 3 and 4, respectively.

3   Research Results

The research results are divided into two parts, with the first part presenting the results of the model development of each classifier. The second part presents the results of model development with majority voting techniques.

3.1 Generated Model Results

Five classification techniques for decision-making are provided in the first step of the research framework. The excellent performance analysis model results with the cross-validation technique and confusion matrix assessment for each predictive classifier are presented in Tables 3 and 4, respectively.

Table 3. Summarizes the results of the five classifiers' performance model analysis

Classifiers	Generalized linear model			Neural Network (NN)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Class dropped	99.13%	99.13%	99.13%	99.12%	98.26%	98.69%
Class scheduled	77.44%	85.81%	81.41%	75.78%	82.43%	78.97%
Class delayed	68.12%	55.29%	61.04%	61.64%	52.94%	56.96%
Accuracy	*82.80%*			*80.42%*		
Classifiers	Decision Tree (DT)			Naïve Bayes (NB)		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Class dropped	99.13%	99.13%	99.13%	98.21%	95.65%	96.91%
Class scheduled	72.09%	83.78%	77.50%	77.05%	63.51%	69.63%
Class delayed	59.02%	42.35%	49.31%	50.88%	68.24%	58.30%
Accuracy	*78.72%*			*75.29%*		
Classifiers	k-Nearest Neighbor (k-NN)					
	Precision		Recall		F1-Score	
Class dropped	90.65%		84.35%		87.39%	
Class scheduled	66.31%		83.78%		74.03%	
Class delayed	42.59%		27.06%		33.09%	
Accuracy	*70.09%*					

Table 3 summarizes the performance analysis results of the five predictive classifiers, which showed that the classifier with the highest accuracy was the Generalized Linear Model (GLMs), with 82.80% accuracy. The second most accurate predictive classifier is the Neural Network (NN), with 80.42% accuracy. The third most accurate predictive classifier is the Decision Tree (DT), with 78.72% accuracy. The top three models with the highest accuracy were computed for majority voting to create a predictive model that re-tested the original data on the cross-validation technique and confusion matrix assessment. The detailed results of the performance model analysis organized by the classifier are presented in Table 4.

Table 4. The results of the performance model analysis organized by the classifier

Generalized Linear Model (GLMs): Accuracy 82.77%

Predicted\Actual	True dropped	True scheduled	True delayed	Class precision
Pred. dropped	114	0	1	99.13%
Pred. scheduled	0	127	37	77.44%
Pred. delayed	1	21	47	68.12%
Class recall	99.13%	85.81%	55.29%	

Neural Network (NN): Accuracy 80.42%

Predicted\Actual	True dropped	True scheduled	True delayed	Class precision
Pred. dropped	113	0	1	99.12%
Pred. scheduled	0	122	39	75.78%
Pred. delayed	2	26	45	61.64%
Class recall	98.26%	82.43%	52.94%	

Decision Tree (DT): Accuracy 78.72%

Predicted\Actual	True dropped	True scheduled	True delayed	Class precision
Pred. dropped	114	0	1	99.13%
Pred. scheduled	0	124	48	72.09%
Pred. delayed	1	24	36	59.02%
Class recall	99.13%	83.78%	42.35%	

Naïve Bayes (NB): Accuracy 75.29%

Predicted\Actual	True dropped	True scheduled	true delayed	Class precision
pred. dropped	110	0	2	98.21%
Pred. scheduled	3	94	25	77.05%
Pred. delayed	2	54	58	50.88%
Class recall	95.65%	63.51%	68.24%	

k-Nearest Neighbor (k-NN): Accuracy 70.09%

Predicted\Actual	True dropped	True scheduled	True delayed	Class precision
Pred. dropped	97	2	8	90.65%
Pred. scheduled	9	124	54	66.31%
Pred. delayed	9	22	23	42.59%
Class recall	84.35%	83.78%	27.06%	

3.2 Majority Voting Prototype Model

After developing and selecting the model with the five classifiers, this section carried out two subsections: The first subsection is ensemble techniques for creating the majority voting models with the top three and all classifiers. The second subsection evaluates the two models' comparative majority voting model performance.

The first subsection started in the second step of the research framework. The second step was calculating the confidence value of each record's prediction with previously selected techniques classified by class to vote. The third step is considering voting to choose a category from each classifier's highest predicted confidence value. In the fourth step, two parts of the majority voting model exist. The first part was a majority voting with the top three most accurate models, and the second was majority voting with all five modeling techniques.

The results of the confident analysis of each classifier in the second and third step calculations and the consequences of two ensemble majority voting models in the fourth step were released as follows: https://bit.ly/3oGWf4l. To conceal the data and prevent compromise on the rights of the informant, the researchers reworked the student code, which made it irreversible or damaging to the person providing the information.

The operating result of the second subsection is a comparison of two majority voting models. A summary of the majority voting for both models was published as follows: https://bit.ly/3oGWf4l. The researchers then compared the majority voting results with the actual data to calculate the efficiency of the two models. The results are summarized in Tables 5 and 6, respectively.

Table 5. The comparison of two majority voting models

Vote ensemble	3-Ensemble			All classifiers		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
class dropped	99.14%	100.00%	99.57%	97.46%	100.00%	98.71%
Class scheduled	78.75%	85.14%	81.82%	79.11%	84.46%	81.70%
Class delayed	69.44%	58.82%	63.69%	68.06%	57.65%	62.42%
Accuracy	*83.62%*			*83.04%*		

Table 5 compares two models using ensemble techniques to select the best predictive model for preventing students' dropout. The researchers found that the 3-ensemble technique model with three classifiers had the highest accuracy, with an accuracy of 83.62%. The three classification techniques consist of Generalized Linear Model (GLMs), Neural Network (NN), and Decision Tree (DT). The model performance of the 3-ensemble classifiers is presented in Table 6.

Table 6 presents the efficacy evaluation of the model. The researchers found that the model could predict the dropout students with 100% accuracy based on the displayed recall values. In addition, the overall model accuracy was high, with an accuracy of 83.62%. The researchers compared their findings with past research and discussed important issues later.

Table 6. The model performance of the 3-ensemble classifiers

3-Ensemble Classifiers: Accuracy 83.62%				
Predicted\Actual	True dropped	True scheduled	True delayed	Class precision
Pred. dropped	115	0	1	99.14%
Pred. scheduled	0	126	34	78.75%
Pred. delayed	0	22	50	69.44%
Class recall	100.00%	85.14%	58.82%	

4   Research Discussion

In this research, researchers studied and developed a predictive model for preventing student dropout at the university level using majority voting and data mining technique. The most rational model of this research was the preventing prediction model with a 3-ensemble majority voting technique, as shown in Table 5. The machine learning tools used as a component of the majority voting model included Generalized Linear Model (GLMs), Neural Network (NN), and Decision Tree (DT). There are interesting findings from this research. From the development of the model, researchers found that the model could predict the results with a high level of accuracy (83.62%). The weak point of this model was that it signified a moderately delayed class, as shown in the F1-Score, which was 63.69%, as shown in Table 5. However, the model predicted dropped and scheduled classes with high accuracy with the F1-Score of 99.57% and 81.82%, as shown in Table 5. Additionally, the model could predict student dropout with 100% accuracy by analyzing the model's performance with cross-validation techniques and confusion matrix assessment, as shown in the Recall value in Table 6.

This research refutes Nuankaew's research [9, 10] by providing a substantial improvement in the original study. Nuankaew's [9] weakness is that it uses only one prediction technique. Nuankaew's [10] weakness is that it doesn't consider the problem of students' delayed graduation. All the weaknesses have been refined and revised to a more excellent quality that the entire process in this research has been presented.

5   Conclusion

The dropout problem among university students is a loss of educational opportunities leading to a shortage of skilled and knowledgeable workers in the labor market. In this research, the main research objectives were to improve the predictive model for preventing dropouts among university students using majority voting and data mining techniques. There are two objectives of the research. The first objective is to study the context and improve the student dropout prevention model at the university level. The second objective is to compare the past university students' dropout models. The data used in this research were students' academic achievements in the Department of Business Computer at the School of Information and Communication Technology,

University of Phayao, during the academic year 2012–2016. There are a total of 254,456 transactions, which have been extracted from the data of 348 students.

The researchers found that the highest number of dropouts in the first-year university were 67 students, representing 58.26%, as shown in Table 2. Therefore, the researchers developed a predictive model for preventing dropout among university students based on course achievement in the first and second semesters of first-year university studies. The model that has been developed uses a combination of majority voting techniques and data mining techniques. The researchers found that the practical model for this research was using 3-ensemble majority voting techniques with a high level of accuracy, with an accuracy of 83.62%, as shown in Table 5. Furthermore, the efficacy evaluation results of the 3-ensemble majority voting model are presented in Table 6. The researchers found that the improved model performed better than Nuankaew's research [9, 10]. This research addresses the weaknesses of all previous research [9, 10], which uses a wider variety of machine learning techniques and controls to prevent students' delayed graduation in higher education.

Based on this research, the researchers would like to suggest guidelines for using the research results as information to solve the problem of student dropout at the university level as follows: (1) educational institutions should focus on and formulate a plan to solve the problem of long-term dropouts through the cooperation of educational institutions and program administrators. (2) Those involved should put the research results into practice to prevent student dropouts at the university level and manage students to complete their studies on time.

6   Limitation

The limitation of this research is that the researcher takes a long time to collect the data, and this is because the program has a four-year study plan and allows students to spend twice the time in their educational program. It may seem that researchers have used outdated data. In fact, these research findings are used in parallel with the current curriculum, effectively helping to prevent student dropouts.

Conflicts of Interest.   The authors declare no conflict of interest.

References

1. Casanova, J.R., Gomes, C.M.A., Bernardo, A.B., Núñez, J.C., Almeida, L.S.: Dimensionality and reliability of a screening instrument for students at-risk of dropping out from Higher Education. Stud. Educ. Eval. **68**, 100957 (2021). https://doi.org/10.1016/j.stueduc.2020.100957

2. Karimi-Haghighi, M., Castillo, C., Hernández-Leo, D.: A causal inference study on the effects of first year workload on the dropout rate of undergraduates. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) Artificial Intelligence in Education: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part I, pp. 15–27. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-11644-5_2

3. Saccaro, A., França, M.T.A.: Stop-out and drop-out: The behavior of the first year withdrawal of students of the Brazilian higher education receiving FIES funding. Int. J. Educ. Dev. **77**, 102221 (2020). https://doi.org/10.1016/j.ijedudev.2020.102221

4. Luo, Y., Zhou, R.Y., Mizunoya, S., Amaro, D.: How various types of disabilities impact children's school attendance and completion – Lessons learned from censuses in eight developing countries. Int. J. Educ. Dev. **77**, 102222 (2020). https://doi.org/10.1016/j.ijedudev.2020.102222

5. Tinto, V.: From theory to action: exploring the institutional conditions for student retention. In: Smart, J.C. (ed.) Higher Education: Handbook of Theory and Research, pp. 51–89. Springer Netherlands, Dordrecht (2010). https://doi.org/10.1007/978-90-481-8598-6_2

6. Burgos, C., Campanario, M.L., de la Peña, D., Lara, J.A., Lizcano, D., Martínez, M.A.: Data mining for modeling students' performance: a tutoring action plan to prevent academic dropout. Comput. Electr. Eng. **66**, 541–556 (2018). https://doi.org/10.1016/j.compeleceng.2017.03.005

7. de Oliveira, C.F., Sobral, S.R., Ferreira, M.J., Moreira, F.: How does learning analytics contribute to prevent students' dropout in higher education: a systematic literature review. Big Data Cogn. Comput. **5**, 64 (2021). https://doi.org/10.3390/bdcc5040064

8. Nuankaew, P., Nuankaew, W., Nasa-ngium, P.: Risk management models for prediction of dropout students in Thailand higher education. Int. J. Innov. Creativity Chang. **15**, 494–517 (2021)

9. Nuankaew, P.: Dropout situation of business computer students, university of Phayao. Int. J. Emerg. Technol. Learn. (iJET) **14**, 115–131 (2019). https://doi.org/10.3991/ijet.v14i19.11177

10. Nuankaew, P., Nuankaew, W., Teeraputon, D., Phanniphong, K., Bussaman, S.: Prediction model of student achievement in business computer disciplines. Int. J. Emerg. Technol. Learn. (iJET). **15**, 160–181 (2020). https://doi.org/10.3991/ijet.v15i20.15273

11. Deng, X., Liu, Q., Deng, Y., Mahadevan, S.: An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inf. Sci. **340–341**, 250–261 (2016). https://doi.org/10.1016/j.ins.2016.01.033