Improving Dropout Forecasting during the COVID-19 Pandemic through Feature Selection and Multilayer Perceptron Neural Network

Sumitra Nuanmeesri, Lap Poomhiran, Shutchapol Chopvitayakun, and Preedawon Kadmateekarun

***Abstract*—Nowadays, online education in universities is mature from the situation of COVID-19 spread. It has greatly changed the learning environment in the classroom and has also resulted in students' dropout. The purpose of this study is to examine the factors that influence students' dropout in the department of information technology of Suan Sunandha Rajabhat University. Students enrolled in the subjects in the COVID-19 pandemic situation must study online at their homes and must take online courses. This study investigated 1,650 student records, 19,450 enrollments, 16,200 grades, and 11,780 social media accounts that had access to all online courses. It was examined how to improve the model's performance by combining feature selection with a multilayer perceptron neural network method. The model was compared to student dropout predictions generated by Logistic regression, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, and Multilayer Perceptron Neural Network, with feature selection (1). The 10-Folds Cross Validation method was used to determine the efficiency of the Gain Ratio, Chi-Square, and Correlation-based Feature Selection models to compare accuracy, precision, sensitivity, F1 score, and classification error rate (e). After adjusting the modeling parameters, the multilayer perceptron neural network method combined with CFS characterization achieved an accuracy of 96.98%. The study's findings indicate that the feature selection technique can be used to improve the neural network model's efficiency in predicting student dropout during a COVID-19 pandemic. Furthermore, the simulation can improve student dropout forecasting during spread out that persists.**

***Index Terms*—Dropout prediction, learning analytics, machine learning, feature selection, MLP.**

I. Introduction

Many problems are being experienced by university students who are using the online learning method amid the COVID-19 outbreak. The main problem that directly affects the university is that students drop out. Although it is an undergraduate degree, students are already at the level of adolescence. Online learning affects learning due to an environment that lacks the stimulation of face-to-face interactions in the actual room learning with lecturers and classmates. E-learning dropout rates generally tend to be higher than face-to-face education [1]. With the spread of the Covid-19 situation, there are still new species continuously, and the number of infected people has not decreased in Thailand [2]. It is quite inevitable that universities still have to study online. Student dropout is one of the most complex issues confronting both students and academic institutions. Students drop out for various reasons, including poor grades, a lack of funds for registration fees, or even resigning from university due to a change in study major or moving to a new higher education institution. The resignation of a student is one of the most complicated and unpleasant developments. Student dropout is a decision that could have a negative impact on the lives of students. From an economic standpoint, it also has a negative impact on academic institutions. This dropout effect may impact educational quality assurance, budgeting, and educational management. It can also be a sign of potential problems in the educational system and the quality of the curriculum. Furthermore, it may influence changes in teaching methods and accommodate learning situations that necessitate social distancing and proper management of teaching material. Predicting dropouts is a solution to the aforementioned issues and a task that can benefit students. Educational data is currently collected in a systematic procedure. It enables teachers to keep track of their student's academic progress, including assessing and predicting the likelihood of student dropout. In the situation of COVID-19 spread, students who have studied in the classroom have changed to fully online learning [3], [4]. It is important to note that recognizing students at risk of dropping. Universities must know the genuine reason for students' dropout and create techniques and alterations strategies to incorporate viable dropout avoidance methodologies. They focus on students' needs to help them prevent dropouts in time, empowering complete to complete their courses on time successfully. Moreover, to find a way to help alleviate education, social, and economic problems in the midst of the COVID-19.

Therefore, the article's primary goal is to improve student dropout forecasting with the feature selection of the available datasets of educational data. This work compares the performance of several machine learning classifiers that can predict student dropout with sufficient accuracy.

The remainder of this article has the following structure. First, related works are discussed in Section II. Then, section III describes the feature selection methodology and machine learning methods used to analyze the dataset collected for dropout forecasting. Next, in Section IV, the authors evaluate the selected models by comparing their performance. Finally, the final work concludes, and further extensions of this work

Manuscript received March 7, 2022; revised April 26, 2022.

Sumitra Nuanmeesri, Shutchapol Chopvitayakun, and Preedawon Kadmateekarun are with the Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok 10300, Thailand (e-mail: Sumitra.nu@ssru.ac.th, shutchapol.ch@ssru.ac.th, preedawan.ka@ssru.ac.th).

Lap Poomhiran is with the Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: lap.p@email.kmutnb.ac.th).

doi: 10.18178/ijiet.2022.12.9.1693

are in Section V.

II. Literature Reviews

Previous research for education prediction in the section of predicts the probability of students completing or failing a course has used a variety of techniques, which were discovered. These techniques were used to analyze, predict, or model, as in the use of correlation rules to predict the probability of students completing or failing a course with learning analytics (LA) [5], [6]. A reliable performance prediction can be utilized to recognize frail understudies and give input to understudies or foresee predict students' failure [7]. The Decision Tree (DT) technique was used to find the relationship between examination of grades accomplished from courses and traits characterizing credits gotten by understudies, and their average grades can moreover be significant [8] to success in learning [9]. They perform two approaches of machine learning, Logistic Regressions (LR) and DT, to predict student dropout [10]. Furthermore, various techniques were used to classify the data to create models to predict student dropout, such as the support vector machine [11]. Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine (SVM) techniques are applied to the task of dropout occurrences prediction for university students. The results show that the prediction accuracy varied between 62% and 87% [12]. Some research studies have compared the machine learning techniques for predicting student dropout in the university. The researchers created a forecasting model using LR, DT, RF, SVM, and Neural Network (NN) algorithms. The results discovered that the prediction accuracy varied between 77% and 93% [1].

Effective expectations depend on utilizing reliable, precise, and critical information [13]. The number of the dataset is maybe a commonplace trap for studies in educational data mining, learning analytics, and educational technology research [14]. It is challenging to reach the volume of educational data that would completely fulfill the machine learning necessities other than utilizing logs [15] and how the discoveries ought to be translated from the instructive point of view [16].

The dimension reduction algorithm simplifies the data by projecting the actual feature space onto a lower dimension while keeping relevant information out of the parameter setting. SVM is a reliable data classification and regression technique. Maximize the margins in the feature area to find the best hyperplane separation. The training data used in the maximization process, known as the RF support vector, generates a collection of tree-structured classifiers that are randomly joined together. It has been used in the literature for a wide range of regression and prediction tasks [17], and feature selection is used in associated with data classification methods to maximize efficiency. As demonstrated by the work that used the Filter Ranker Method or Correlation-based Feature Selection (CFS) to select key features associated with data classification [18]-[20].

On the other hand, while mentioning the problem, most of the previous works do not focus on selecting data features for the best performance for predicting student dropout and an excellent survey of different techniques for the students' fail and success prediction.

The preceding research investigated factors or attributes used in an analysis that differ from context to context and available data sets. The data and sample groups have a wide range of effects on dropout rates. Certain factors may have a different impact on forecast accuracy. In each context, the performance of techniques for classifying or predicting varies. As a result, in this study, the feature selection method was used to select important factors for student dropout, which was then combined with data mining techniques to predict student dropout of Suan Sunandha Rajabhat University students majoring in Information technology. To face the challenge of anticipating dropouts for new and current students who may drop out, taking a data-driven approach with automated learning processes to obtain a model that can collect relevant information about the specific context in which the dropout can occur in the spreading situation of COVID-19 that may persist and under the normal circumstances.

III. Methodology

A. Data Collection and Preparation

The student dropout forecast was based on 1,650 student records, 19,450 enrollments, 16,200 grades, and 11,780 social media accounts that accessed all the online courses. Suan Sunandha Rajabhat University has established regulations on student dropout that are consistent with the curriculum structure to determine the factors influencing student dropout (Suan Sunandha Rajabhat University, 2017). This study chose data by removing redundant information aimed to find out factors derived from the original data to filter the data to eliminate inaccurate information. There are 2,400 data items available for further processing. Data transformation is carried out as follows by substituting and transforming all data into a nominal format. Next, the authors transform the data into a format that can be analyzed and divide the target classes into studying and graduation. Finally, the data has been converted to a total of 1,225 items remain (out of a total of 2,400 items). The data preparation results in residual attributes, which have sixteen attributes, as shown in Table I.

TABLE I: List of Factors Used in Data Analysis

Attribute	Description	Possible values
Schoolold	Former school	school=school, collage = college, nfe = Non-Formal Education
Schoolmajor	High school major	science = Science/Math, social = Arts
Parentcareer	Parent's career	unknown = unknown none = no career owner = business owner agriculture = farmer/fisherman freelance = freelancer govern = government officer employee = employee gemployee = government employee semployee = state enterprise employee etc = Other

Parentincome	Parent's income (Bath per year)	unknown = unknown none = no income low = not over 100,000 medium = 100,001-200,000 high = over 200,000
Sibling	Number of siblings	none = None one = 1 sibling two = 2 siblings three = 3 siblings over = more than 3 siblings
Loan	Student loan status	yes = loaned, no = no loan
Dropsubject	Withdrawal of course (W)	yes = Attain grade W before no = Never got grade W
Droptemp	Maintaining student status	yes = Used to maintain no = Never maintain
Internetpay	Internet service	monthly = post-paid/subscribed top up = pre-paid
Socialclass	Joining a study group through social media	yes = join no = not to join
GPA	Cumulative grade point average	low = 0.00-1.99 mid = 2.00-2.99 high = 3.00-4.00
GPAold	High school cumulative grade point average	low = 0.00-1.99 mid = 2.00-2.99 high = 3.00-4.00
GPAfac	Average grades of subjects in the faculty	low = 0.00-1.99 mid = 2.00-2.99 high = 3.00-4.00
GPAnone	Average grades for non-faculty subjects	low = 0.00-1.99 mid = 2.00-2.99 high = 3.00-4.00
GPAflang	Average grades for foreign language courses	low = 0.00-1.99 mid = 2.00-2.99 hi = 3.00-4.00
Graduate	Student status	yes = graduate no = out of student status

B. *Forecasting Model Development*

The authors used weight analysis techniques including Gain Ratio (GR), Chi-Square (CS), and CFS to select key features, as well as modeling to compare the model's efficiency, including LR, DT, RF, Naive Bayes (NB), SVM, Multilayer Perceptron Neural Network (MLP) using the Weka software for modeling. The process of modeling for forecasting student dropout is as follows:

1) The LR, DT, RF, NB, SVM, and MLP models were created from a total of sixteen attributes (Table I), where the last attribute is class output. Their performance was compared to accuracy, precision, sensitivity, F1 score, and classification error rate (e).
2) The feature selection techniques, including GR, CS, and CFS methods, were used to select factors. These feature selection methods will select only the expected attributes directly affecting the model's efficiency. Occasionally, high-accuracy models are subject to discrepancies, or the classification error rate is also high. This high error rate may cause by all attributes used in the modeling, and some attributes may not have or have little effect on the model's efficacy. For this reason, it can be a decreased performance when the model is put into production. Thus, the sample data of 1,225 items were processed by GR, CS, and CFS methods. Then, the remaining factors or attributes become features with classifying capabilities. Finally, they are selected and used to develop a more efficient model.

The single-attribute evaluator technique is used for the GR and CS method with the attribute ranking approach which is far faster. All attributes were ranked based on the average metric of all sample data. The subset attribute evaluator was applied for the CFS method that ranked all sample data. However, in the experimental process, the number of attributes was gradually reduced to determine how many retaining attributes were appropriate for the data used in this research.

3) The LR, DT, RF, NB, SVM, and MLP models were applied to the selected attributes to take into account the number of attributes rated by 50 GR and CS fraction of all attributes and an amount equal to the CFS yield greater than 0%. The accuracy, precision, sensitivity, F1 score, and classification error rate were used to compare efficiency.
4) In order to increase the model's effectiveness in predicting student dropout, adjust the modeling method parameters from the previous step with the highest accuracy. The performance is then compared to accuracy, precision, sensitivity, F1 score, and classification error rate.

The process of forecasting the student dropout is shown in Fig. 1.

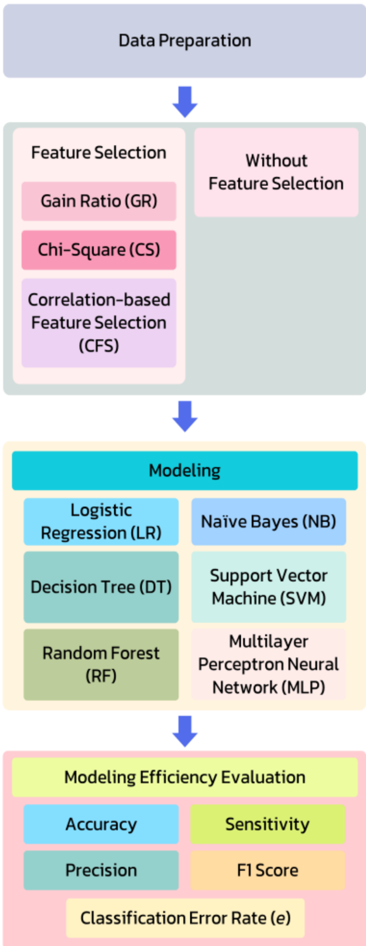


Fig. 1. The process of forecasting student dropout.

C. *Efficiency Evaluation*

Model performance is evaluated using the K-Fold

Cross-Validation (10-Folds) method, which compares the accuracy, precision, sensitivity [21], [22], F1 score, classification error rate [1] as in (1) to (5), which are developed by the LR, DT, RF, NB, SVM, and MLP techniques. These models are combined with feature selection using the GR, CS, and CFS.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ score = \sqrt{Recall \times Precision} \quad (4)$$

$$e = \frac{FP+FN}{TP+TN+FP+FN} \quad (5)$$

IV. Results

The following are the results of the study comparing model performance and feature selection:

A. The Results of the Model's Efficacy of the Non-factor Model

This research compares the efficiency of models generated with the LR, DT, RF, NB, SVM, and MLP algorithms from a total of sixteen attributes (without feature selection), where the Graduate attribute was defined as the target class. It was discovered that the RF algorithm (C4.5) produced the highest accuracy results was 88.33%, followed by LR, SVM, MLP, DT, and NB with an accuracy of 87.59%, 85.55%, 83.92%, 80.00%, and 77.14%, respectively. While the classification error rate showed that the MLP algorithm had the lowest value was 8.49%. Other algorithms followed by SVM, LR, DT, RF, and NB with error rates were 15.35%, 15.51%, 15.59%, 16.49%, and 19.35%, respectively. The efficacy results of each model without the feature selection process are shown in Fig. 2.

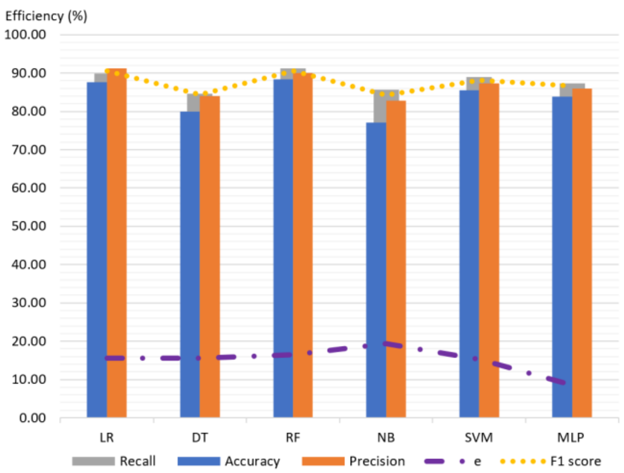


Fig. 2. The efficacy results of the non-factor model.

B. The Results of the Feature Selection

The results of GR, CS, and CFS methods showed that the number of features selected by the GR and CS methods gave almost the same results. However, both methods differ in precedence between attribute 6th and 7th, which alternate between Loan and Parentincome. Nevertheless, the remaining attribute priorities are the same in which the top five are GPA, GPAnone, Socialclass, GPAfac, and Internetpay. While The CFS method yields attributes over 0% weighted in seven attributes: GPA (100%), GPAnone (100%), Socialclass (95%), GPAfac (75%), Loan (70%), Internetpay. (5%), and Dropsubject (5%). The results of the attribute priorities were ranked and selected using the three feature selection methods shown in Table II.

TABLE II: The Results of Feature Selection Methods

Rank	GR	CS	CFS (% of weight)
1st	GPA	GPA	GPA (100%)
2nd	GPAnone	GPAnone	GPAnone (100%)
3rd	Socialclass	Socialclass	Socialclass (95%)
4th	GPAfac	GPAfac	GPAfac (75%)
5th	Internetpay	Internetpay	Loan (70%)
6th	Loan	Parentincome	Internetpay (5%)
7th	Parentincome	Loan	Dropsubject (5%)
8th	Dropsubject	Dropsubject	Droptemp
9th	Droptemp	Droptemp	Parentincome
10th	GPAflang	GPAflang	GPAflang
11th	GPAold	GPAold	GPAold
12th	Sibling	Sibling	Sibling
13th	Schoolmajor	Schoolmajor	Schoolmajor
14th	Parentcareer	Parentcareer	Parentcareer
15th	Schoolold	Schoolold	Schoolold

According to Table II, the attributes that might directly affect the model's efficacy could be the first seven attributes used to develop the model. Therefore, the top five attributes and the top seven attributes were selected and used to develop models for forecasting student dropout.

C. The Results of the Model's Efficacy in Combination with the Feature Selection

The results showed that the top five features have better model performance than the top seven features for forecasting student dropout modeling. In addition, the overall efficiency was increased from around 0.33% to 0.82% for the LR, DT, RF, and NB algorithms. As for MLP, there was an efficiency increase of around 6%. Additionally, the CFS method provides better model performance values than the GR or CS methods. When comparing the performance of models using the top five attributes (atts) with different algorithms, the CFS-qualified MLP model had the highest accuracy of 91.48% and the lowest classification error rate of 4.08%. The following models were RF, LR, SVM, DT, and NB with 90.04%, 88.82%, 87.10%, 86.61%, and 79.18% accuracy, respectively. The results of the model's efficacy, including accuracy (Acc), precision (Prec), recall, F1 score, and classification error rate (e) are shown in Table III.

TABLE III: THE EFFICACY OF MODEL WITH FEATURE SELECTION

Models	Acc	Prec	Recall	F1	e
LR+GR (7 atts)	87.59	91.28	89.94	90.61	5.51
LR+CS (7 atts)	87.59	91.28	89.94	90.61	15.51
LR+GR (5 atts)	88.41	91.96	90.61	91.28	14.94
LR+CS (5 atts)	88.41	91.96	90.61	91.28	14.94
LR+CFS (5 atts)	88.82	92.35	90.89	91.61	13.96
DT+GR (7 atts)	86.29	93.77	86.44	90.03	9.14
DT+CS (7 atts)	86.29	93.77	86.44	90.03	9.14
DT+GR (5 atts)	86.61	94.39	86.06	90.13	8.98
DT+CS (5 atts)	86.61	94.39	86.06	90.13	8.98
DT+CFS (5 atts)	83.10	87.31	86.08	86.69	12.90
RF+GR (7 atts)	88.33	89.99	91.29	90.64	16.49
RF+CS (7 atts)	88.33	89.99	91.29	90.64	16.49
RF+GR (5 atts)	89.14	90.83	91.90	91.36	15.59
RF+CS (5 atts)	89.14	90.83	91.90	91.36	15.59
RF+CFS (5 atts)	90.04	91.97	92.21	92.09	14.61
NB+GR (7 atts)	77.14	82.82	85.67	84.23	19.35
NB+CS (7 atts)	77.14	82.82	85.67	84.23	19.35
NB+GR (5 atts)	77.96	83.46	86.32	84.88	18.69
NB+CS (5 atts)	77.96	83.46	86.32	84.88	18.69
NB+CFS (5 atts)	79.18	84.84	86.84	85.83	17.22
SVM+GR (7 atts)	85.55	87.38	88.92	88.15	15.35
SVM+CS (7 atts)	85.55	87.38	88.92	88.15	15.35
SVM+GR (5 atts)	85.96	87.78	89.42	88.60	12.49
SVM+CS (5 atts)	85.96	87.78	89.42	88.60	12.49
SVM+CFS (5 atts)	87.10	88.85	90.27	89.55	11.02
MLP+GR (7 atts)	83.92	85.98	87.28	86.63	8.49
MLP+CS (7 atts)	83.92	85.98	87.28	86.63	8.49
MLP+GR (5 atts)	90.61	92.85	92.25	92.55	4.49
MLP+CS (5 atts)	90.61	92.85	92.25	92.55	4.49
MLP+CFS (5 atts)	91.84	93.53	93.53	93.53	4.08

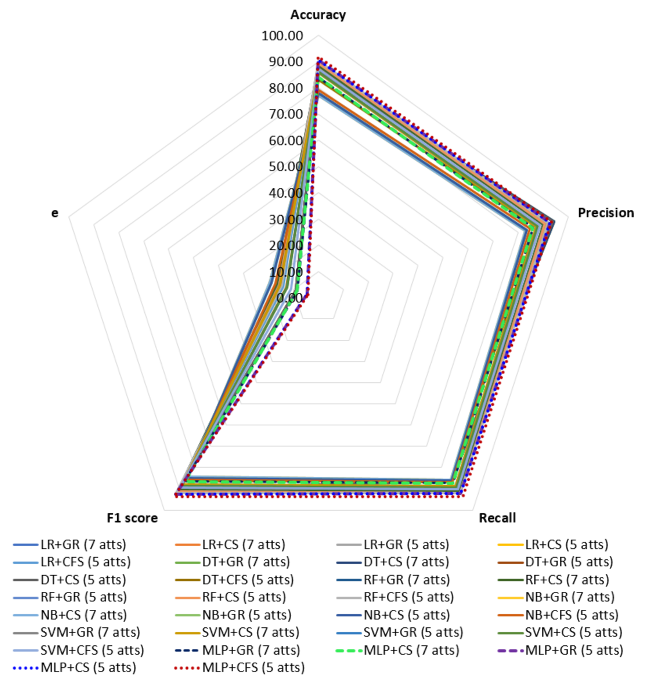
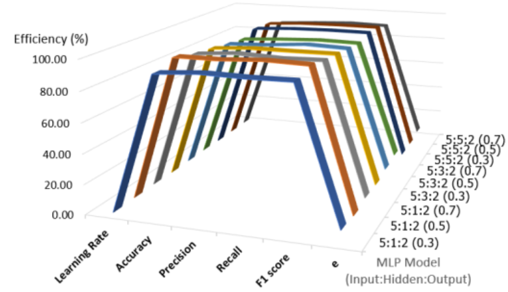


Fig. 3. The results of the model's efficacy in combination with the attribute selection.

The comparison of the performance of models generated using the LR, DT, RF, NB, SVM, and MLP algorithms are based on the feature selection for five and seven attributes, as shown in Fig. 3.

D. The Results of CFS Qualification Combined with the MLP Model's Efficacy

Based on the experimental results from the previous step, the authors adjusted the modeling parameters using the MLP method to test the model's effectiveness by changing the learning rate to 0.3, 0.5, and 0.7, the number of MLP hidden layer nodes to 1, 3, and 5, the momentum to 0.2, the input layer to 5, and the output layer to 2. According to the results, the model with the highest accuracy was created with input layer (I) = 5, hidden layer (H) = 1, output layer (O) = 2, learning rate = 0.5, and momentum = 0.2, as shown in Fig. 4.



	Learning Rate	Accuracy	Precision	Recall	F1 score	e
5:1:2 (0.3)	0.3	92.24	94.27	93.67	93.97	7.76
5:1:2 (0.5)	0.5	96.98	96.92	98.50	97.70	6.29
5:1:2 (0.7)	0.7	93.39	95.53	94.21	94.87	8.16
5:3:2 (0.3)	0.3	92.00	93.85	93.61	93.73	7.67
5:3:2 (0.5)	0.5	90.04	91.06	93.20	92.13	9.80
5:3:2 (0.7)	0.7	88.57	90.26	91.45	90.85	11.76
5:5:2 (0.3)	0.3	92.16	93.61	94.09	93.85	4.08
5:5:2 (0.5)	0.5	91.59	93.20	93.56	93.38	4.33
5:5:2 (0.7)	0.7	87.84	89.56	90.86	90.21	6.53

Fig. 4. The results of CFS qualification combined with the MLP model's efficacy.

V. CONCLUSION AND DISCUSSION

This study aims to improve the model for predicting student dropout rates in the Information Technology program at Suan Sunandha Rajabhat University by analyzing the factors influencing student dropout during the COVID situation in which students must study online at their residence. Using 1,650 student records, 19,450 enrollments, 16,200 grades, and 11,780 social media accounts that accessed all the online courses, with the method used to create Logistic Regression, Decision Tree, Random Forest, Naive Bayes, Support Vector Machine, and Multilayer Perceptron Neural Network, were used, with feature selection methods including Gain Ratio, Chi-Square, and Correlation-based Feature Selection. The study revealed that the model's accuracy increased when the attribute was reduced to five attributes and used in conjunction with the CFS selection with the highest accuracy. When CFS was combined with MLP, the student dropout prediction accuracy was 91.84%. On the other hand, a learning rate adjustment of 0.5, was the best value that increased the model's accuracy to 96.98%. The CFS method provided the highest level of accuracy in attribute selection. According to the research findings, the following factors influence university student dropout: cumulative grade point average (GPA), cumulative grade point average for non-faculty subjects (GPAnone), and the joining social media group in subjects (Socialclass). It was also discovered that the cumulative grade point average of faculty subjects (GPAfac) and Loan is a secondary factor influencing student dropout.

The findings indicate that feature selection techniques can be used to improve model performance in predicting student dropout when comparing the results between models that were not subject to the feature selection process and those that were processed through GR, CS, and CFS. Furthermore, the CFS combined with MLP model can improve the accuracy of the neural network model, which can then be used to develop a system for predicting university dropouts in the future. This research can examine trends in student retention status and provide helpful advice when an academic institute is faced with critical situations during the COVID-19 Pandemic. Academic performance in any subject or lack of participation in the course group throughout social networks. There is a chance that students will be able to leave midway. Then, specific support services must be provided, such as extended support and the reduction in tuition fees for students to enable students to pay for the cost of internet services used in their distance studies. Personal advice techniques can be arranged for those student groups. The study intends to model an effective way to reduce dropout rates in the coming years to avoid this phenomenon among university students at the start of their studies and keep the student retention rate each year close to that of first-time students. This forecasting model can be used as a strategy to improve the model's performance, which can be helpful for courses from multiple disciplines at the same university or different universities with similar courses. This can be combined with an appropriate definition of course similarity measures. Finally, this forecasting model can be expanded to optimize and facilitate online teaching and learning activities in Learning Management Systems (LMS) or Virtual Learning Environments (VLEs) like Moodle and Google Classroom [12].

This is especially important when using classifiers that do not include gender, race, or religion concerning the attributes in the model. In addition, Deep Learning-based models will be developed in the future to expand the multilayer analysis of each layer and to incorporate visual research to aid model forecasting.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The authors had conducted the research and finished this paper cooperatively; all the Authors had approved the final version.

Acknowledgment

The authors are grateful to the Institute for Research and Development, Suan Sunandha Rajabhat University, and the Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, who supported and gave this research opportunity.

References

[1] J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses using different machine learning techniques," *Applied Sciences*, vol. 11, no. 3130, pp. 1-19, 2021.

[2] World Health Organization, *Coronavirus disease 2019 (COVID-19) WHO Thailand Situation Report 219 - 19 January 2022*, 2022.

[3] M. Hassan, "Online teaching challenges during COVID-19 pandemic," *International Journal of Information and Education Technology*, vol. 11, no. 1, pp. 41-46, 2021.

[4] S. Smith, D. Cobham, and K. Jacques, "The use of data mining and automated social networking tools in virtual learning environments to improve student engagement in higher education," *International Journal of Information and Education Technology*, vol. 12, no. 4, pp. 263-271, 2022.

[5] A. Alamri, M. AlShehri, A. Cristea, F.D. Pereira, E.Oliveira, L. Shi, and C. Stewart, "Predicting MOOCs dropout using only two easily obtainable features from the first week's activities," in *Proc. the International Conference on Intelligent Tutoring Systems*, Kingston, Jamaica, 3-7 June 2019, Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2019, vol. 11528, pp. 163-173.

[6] C. Romero, S. Ventura, R. Baker, and M. Pechenizkiy, *Handbook of Educational Data Mining (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, CRC Press, Boca Raton, FL, USA, 2010.

[7] J. Skalka and M. Drlik, "Automated assessment and microlearning units as predictors of at-risk students and students' outcomes in the introductory programming courses," *Applied Science*, vol. 10, no. 4566, 2020.

[8] C. Lang, G. Siemens, A. Wise, and D. Gasevic, *Handbook of Learning Analytics*, SOLAR: Beaumont, AB, Canada, 2017.

[9] S. Nagrecha, J. Z. Dillon, and N. V. Chawla, "MOOC dropout prediction: Lessons learned from making pipelines interpretable," in *Proc. the 26th International Conference on World Wide Web Companion*, pp. 351-359, 2017.

[10] L. Kemper, G. Vorhoff, and B. U. Wigger, "Predicting student dropout: A machine learning approach," *European Journal of Higher Education*, vol. 10, no. 1, pp. 28-47, 2020.

[11] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in *Proc. EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pp. 60–65, 2014.

[12] F. D. Bonifro, M. Gabbrielli, G. Lisanti, and S. P. Zingaro, "Student dropout prediction," *Lecture Notes in Computer Science*, vol. 12163, 2020.

[13] C. Jin, "MOOC student dropout prediction model based on learning behavior features and parameter optimization," *Interactive Learning Environments*, 2020.

[14] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educational Technology Research and Development*, vol. 68, pp. 1961-1990, 2020.

[15] A. A. Mubarak, H. Cao, and W. Zhang, "Prediction of students' early dropout based on their interaction logs in online learning environment," *Interactive Learning Environments*, 2020.

[16] M. Drlik and M. Munk, "Understanding time-based trends in stakeholders' choice of learning activity type using predictive models," *IEEE Access*, vol. 7, pp. 3106-3121, 2018.

[17] C. Iamprik and K. Sudadet, "Predicting student's data for performance by data mining," in *Proc. The Thirteenth National Conference on Computing and Information Technology NCCIT2017*, pp. 32-37, 2017.

[18] P. Rawengwan and P. Seresangtakul, "Model for forecasting educational status of students (in Thai)," in *Proc. the National and International Graduate Research Conference*, pp. 273-283, 2017.

[19] S. Sittichat, "Study of educational attributes using data mining technique (in Thai)," *Information Technology Journal*, vol. 13, no. 2, pp. 20-28, 2017.

[20] C. Boonprasom and C. Sanrach, "Predictive analytic for student dropout in undergraduate using data mining technique," *Technical Education Journal King Mongkut's University of Technology North Bangkok*, vol. 9, no. 1, 142-151, 2018.

[21] S. Nuanmeesri, "Mobile application for the purpose of marketing, product distribution and location-based logistics for elderly farmers," *Applied Computing and Informatics*, 2019.

[22] S. Nuanmeesri, "Development of community tourism enhancement in emerging cities using gamification and adaptive tourism recommendation," *Journal of King Saud University — Computer and Information Sciences*, 2021.

Sumitra Nuanmeesri received the Ph.D. in information technology at King Mongkut's University of Technology North Bangkok, Thailand. She is an assistant professor in Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand. Her research interests include speech recognition, data mining, deep learning, image processing, web and mobile application, supply chain management system, augmented reality (AR) and virtual reality (VR) development, robotics, education technology, and the internet of things (IoT).

Lap Poomhiran has been teaching in Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand. He is currently a Ph.D. student in information technology, Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. His research interests include web and mobile programming, augmented reality (AR) and virtual reality (VR) development, image processing, data mining, machine learning, deep learning, education technology, and the internet of things (IoT).

Shutchapol Chopvitayakun has been teaching in Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand. His research interests include data mining, web and mobile application, the internet of things (IoT), and education technology.

Preedawon Kadmateekarun received the Ph.D. in information technology at King Mongkut's University of Technology North Bangkok, Thailand. She is an assistant professor in Department of Information Technology, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Thailand. Her research interests include data mining, web and mobile application, augmented reality (AR) and education technology.