International Journal of Applied Science and Engineering

Special Issue-12th International Conference on Awareness Science and Technology (iCAST 2023)

Why first-year e-students are dissatisfied: Machine learning methods for enhancing retention

Thao-Trang Huynh-Cam [1], Long-Sheng Chen [1]*, Van-Canh Nguyen [2], Thanh-Huy Nguyen [3], Tzu-Chuen Lu [1]

[1] *Department of Information Management, Chaoyang University of Technology, Taichung City 413310, Taiwan*
[2] *Quality Assurance Office, Dong Thap University 783, Cao Lanh City, Dong Thap Province, Vietnam*
[3] *Foreign Language Faculty, Dong Thap University 783, Cao Lanh City, Dong Thap Province, Vietnam*

ABSTRACT

Due to advances in wireless technology, e-learning programs and e-courses have increasingly been employed as a mainstream educational mechanism and potentially become crucial incomes of higher education (HE) worldwide. In practice, e-students are required to have highly qualified e-learning programs and satisfied services. It is also extremely difficult for HE to maintain e-students retention as e-students, especially first-year e-students, easily exit from their e-learning programs or shift from one HE to another HE owing to dissatisfaction. However, the dissatisfaction of first-year e-students has gained limited theoretical and practical attention. Thus, it is essential to explore what features make first-year e-students dissatisfied so that HE may have enough time to issue preventive strategies at the early stages for sustainable e-learning adoption. Thus, this study aimed to extract important features using machine learning methods. Data was obtained by using a 5-point Likert e-questionnaire between May and June 2022, generating 499 valid responses from first-year e-students in a Vietnamese public university. The results showed that DT (90.4%) was superior to SVM (88.8%), LR (88.8%), and MLP (85.0%). The most important features included "easy access e-courses via the school e-learning platform", "adequate personal internet skills", "feeling stimulated to attend e-courses, "stable and uninterrupted e-learning platform", "adequate personal digital devices", "teachers' great efforts to improve students' learning", and "timely responses provisions to students' inquiries". The findings of this study are expected to assist HE policy-makers in minimizing e-students' dissatisfaction and maximizing their satisfaction in order to enhance e-student recruitment and retention, and enhance the quality of e-educational programs.

Keywords: E-student retention, First-year e-student dissatisfaction, Important features for e-student dissatisfaction, Machine learning methods, Vietnam.

OPEN ACCESS

Received: December 27, 2023
Revised: February 19, 2024
Accepted: March 12, 2024

Corresponding Author:
Long-Sheng Chen
lschen@cyut.edu.tw

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

Owing to advances in wireless technology, electronic educational (e-educational) programs and electronic courses (e-courses) have been currently utilized as a mainstream educational mechanism worldwide to satisfy the rising demands of learners (De Melo Pereira et al., 2015; Bossman and Agyei, 2022). Additionally, they have been considered a newly emerging market and significantly increase potential profits of higher education (HE) (Ho et al., 2021; Schulz, 2023). Nevertheless, there is very competitive competition in student recruitment due to the huge number of universities (Yoke, 2018; Ullah et al., 2019). HE has continuously endeavoured to be superior in the competition on designing e-courses and e-educational programs. They must take steps to attract more electronic students (e-students) not only for their financial benefits but also for increasing the prestigious and academic quality.

E-students refer to students registering e-courses offered via the Internet and technological platforms (Muljana and Luo, 2019). E-students are frequently considered themselves as HE customers; hence, unfavourable experiences in learning programs and related facilities represent service dissatisfaction (Yoke, 2018; Jameel et al., 2021). In practice, e-students have required highly qualified e-learning programs and satisfied services. Student dissatisfaction significantly influences student withdrawal decisions (Nevill and Rhodes, 2004; Yoke, 2018). In addition, satisfaction and dissatisfaction are tied to student performances (Yang et al., 2013; Baber, 2020); influence the decision to take or refuse additional e-courses or e-educational programs (Anderson and Srinivasan, 2003; Pham et al., 2019); and are likely to recommend or unrecommend e-courses and e-educational programs to the others in the societies (Perez-Perez et al., 2020; Jameel et al., 2021).

It is extremely difficult for HE to retain e-students retention as e-students, especially first-year e-students, easily exit from their e-educational programs or shift from one HE to another HE owing to dissatisfaction (Violante and Vezzetti, 2015; Johnson et al., 2021). Similar to other HE worldwide, Vietnamese HE, especially in remote regions in Mekong Delta, has recently addressed effective methods for enhancing e-student satisfaction to better e-students recruitment and retention (Pham et al., 2020; Huynh-Cam et al., 2023). It is vital to explore which features influence first-year e-students' dissatisfaction. If HE understands features contributing to the dissatisfaction, it is able to minimize these influenced features to facilitate the greater rate of retention and to prevent withdrawal decisions caused by dissatisfaction.

In surveying studies on important features for e-student satisfaction and dissatisfaction in HE, many prior studies have concluded that e-student dissatisfaction is associated with availability of the e-learning platforms, poor teacher attitude, poor teacher communication skills, and poor teacher professional skills (Fu, 2010); assessment, testing, and final exams (Martín Rodríguez et al., 2019); e-classroom involvement requirements (Skrbinjek and Dermol, 2019); access, attentiveness, communication, and availability (Douglas et al., 2015); students' social perceptive, study-personal life balance, workload, assessments, financial difficulties, and learning environment (Nevill and Rhodes, 2004); and lecturers (Elliot, 2003). From the prior related works surveyed in our research, first-year e-student dissatisfaction, which highly leads to potentials of withdrawals, has gained limited attention in theories and practices. Additionally, many researches emphasized e-student satisfaction in big cities such as Wang et al. (2023) and Lu et al. (2020). Less studies focussed on first-year e-student dissatisfaction and major features influencing their dissatisfaction in Vietnamese rural areas.

In seeking to Machine Learning (ML) methods which is capable to analyse data with a high level of accuracy and to retrieve useful knowledge in a short time (Alnagar, 2020) for exploring important features, ML methods have widely and successfully utilized diverse domains such as farming products (Amkor et al., 2024), social media (Chang et al., 2020; Chen et al., 2021; Kumar and Yadav, 2023), and excessive workloads in workplaces (Sung, 2022). In recent decades, they have been increasingly applied for e-student satisfaction and dissatisfaction (Chen and Su, 2008; Sandiwarno et al., 2023) worldwide. For e-student satisfaction, artificial neural network successfully applied in a public Saudi Arabia university with an accuracy of 92.2% and AUC (area under ROC curve) of 99% (Alnagar, 2020). Decision trees (DT) highly forecasted satisfaction of e-students of a private HE in Slovenia (Skrbinjek and Dermol, 2019). Multilayer perceptron (MLP), random forests (RF), K-nearest neighbour (KNN), support vector machines (SVM), and naïve bayes (NB) were successfully used in a Vietnamese National Academy (Lu et al., 2020; Huynh-Cam et al., 2021). Logistic regression (LR), KNN, SVM, DT, multinomial naive bayes (MNB), gradient boosted decision trees (GBDT), convolutional neural network (CNN), long short-term memory (LSTM), CNN+LSTM, E-learning users' satisfaction detection (El-USD) were utilized to predict Indonesian e-student satisfaction (Sandiwarno et al., 2023). ML was also combined with descriptive methods for e-student satisfaction in University of Jordan and was found to be better than descriptive (Masadeh et al., 2023). For student dissatisfaction, Ullah et al. (2019) combined NB, LR, and RF with descriptive statistics and logistic regression analysis to study the linkage between dissatisfaction and retention in Islamic University Chittagong, Bangladesh. From prior related studies surveyed in this work, limited research employed ML methods for first-year e-student dissatisfaction and enhancing e-student retention and recruitment.

The research gap becomes evident in the following aspects. Firstly, despite existing studies on e-student satisfaction and dissatisfaction, there is limited attention given to the dissatisfaction of first-year e-students. This gap suggests a lack of comprehensive understanding and exploration of factors specifically contributing to dissatisfaction among this particular group. Secondly, many prior studies have emphasized e-student satisfaction in urban settings (i.e. big cities), while less attention has been given to the context of rural areas. The research intends to address this geographical gap by focusing on first-year e-student dissatisfaction in a remote public university in Vietnam. This indicates a need for more localized and context-specific investigations. Thirdly, there is an application of ML methods to analyse data related to e-student dissatisfaction. Although there is a growing trend in utilizing ML for e-student satisfaction and dissatisfaction globally, the research suggests that there is limited emphasis on first-year e-student dissatisfaction specifically. The utilization of ML algorithms such as LR, MLP, SVM, and DT, for this purpose is a notable approach.

The methodology adopted in this work enables HE to survey important features associated with first-year e-student dissatisfaction. The outcomes of this survey may act as an early warning system for HE actions to minimize e-students' dissatisfaction and maximize their satisfaction, enhance retention, enhance recruitment, and enhance the quality of e-educational programs.

2. MATERIALS AND METHODS

2.1 Implementation Procedure

The present case research adopted a quantitative method to forecast first-year e-students' dissatisfaction by using supervised ML algorithms. Fig. 1 depicts the 5-step procedure of implementation.

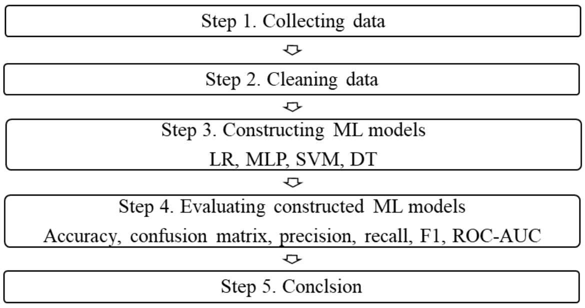


Fig. 1. Implementation procedure

- Step 1. Collecting data

Data was obtained by using an electronic questionnaire (EQ), which had two main parts. Part 1 was used for collecting the basic demographic information of participants. Part 2, including 17 items, was used for constructing ML models. The EQ was distributed to the first-year e-students via the school registered system between May and June 2022 and administered by the Quality Assurance Office of the Vietnamese target university. Thus, there were no duplicated responses and unauthorized respondents. In addition, participants' identity remained anonymous for ethical reasons, higher response percentage, and biased decrease. Participants were asked to rank the level of agreement on all items based on the 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. The time response was approximately 5 min.

- Step 2. Cleaning data

Before constructing ML models, the Microsoft Excel 2016 software was used to transform the original dataset into a useful dataset. In this step, all examples containing missing values were eliminated. The output class "satisfaction" (SAT) was employed as the class label for data distribution. Firstly, number 1 (strongly disagree) combined with number 2 (disagree) to create a new class "disagree"; number 4 (agree) combined with number 5 (strongly agree) to create a new class 'agree" since they were closest. After combining, there were three classes: "agree", "disagree", and "neutral". After that, the classes "agree" and "disagree" were transformed into "satisfied" and "dissatisfied". Since this work only emphasized satisfied and dissatisfied e-students, the class "neutral" anchored at 3, was removed. We also generated features from the content in 17 questionnaire items for ease of interpretation. In addition, we employed Cronbach's Alpha to assess the internal consistency of the measurement tool. Table 1 describes all these generated features and other validity and reliability values. Generally, a threshold of 0.7 is considered acceptable. From the table, it can be observed that the Cronbach's alpha values for the four dimensions (D1–D4) all exceed 0.7, indicating a high reliability of the questionnaire (measurement tool). This means that the questionnaire consistently produces reliable results across multiple users.

- Step 3. Constructing ML models

The present work constructed PMs by using four supervised ML algorithms: LR, MLP, SVM, and DT on Jupyter notebook tool in Python language, which is available at https://jupyter.org/ and https://scikit-learn.org/stable. Each ML model was constructed using a 5-fold cross validation (CV) process (Browne, 2000) with five different training-testing data (80 : 20). The mean value and standard deviation (SD) of the 5-fold CV for every model were utilized for benchmarking prediction performance among four ML models.

As this research focused on the minority class "dissatisfaction", the minority prediction is prior for analysis. For achieving the best accuracy prediction and tackling the imbalanced class problems, this study conducted three cases of class distribution. Fig. 2 shows the data distribution of three classification cases. Case 1 did not apply any rebalanced method. Vice versa, case 2 under-sampled the majority class "satisfaction" and case 3 over-sampled the minority class "dissatisfaction" for rebalancing data. After the best ML model was selected, the Gini index was employed to extract important features for first-year e-student satisfaction and dissatisfaction.

- Step 4. Evaluating constructed ML models

The employed model evaluation methods comprised accuracy, confusion matrix, F1_score, ROC curve (receiver operating characteristic curve), and AUC.

- Step 5. Conclusion

After benchmarking prediction performance among four constructed ML models, we selected the model with the highest performance to retrieve the top-ranking important features for first-year e-students' satisfaction. The present work also offered several suggested measures for maximizing satisfaction and minimizing dissatisfaction among first-year e-students for enhancing retention percentages and higher recruitment rates in terms of sustainable development.

Table 1. Feature description

Feature no.	Feature description (Generated from the questionnaire items)	Mean	SD	Corrected item-total correlation	Cronbach's alpha
D1. E-course organization					0.77
F1	Well-organized e-courses	4.1	0.64	0.62	
F2	User-friendly e-courses	4.1	0.69	0.66	
F3	Updated information/knowledge provided	4.1	0.67	0.54	
D2. Infrastructure and technology					0.78
F4	Smoothly login the required e-learning platform	4.1	0.76	0.63	
F5	Easily access e-courses via the school e-learning platform	4.0	0.76	0.68	
F6	Stable and uninterrupted e-learning platform	3.3	0.99	0.52	
F7	E-learning platform can be synthesized in multi devices	4.1	0.71	0.56	
D3. Teaching staff					0.87
F8	Good professional knowledge of teaching staff	4.3	0.59	0.74	
F9	Variety of teaching techniques	4.2	0.65	0.70	
F10	Enthusiastic and friendly teachers	4.2	0.71	0.76	
F11	Teachers' great efforts to improve students' learning	4.1	0.70	0.74	
F12	Providing timely responses to students' inquiries	4.1	0.71	0.59	
D4. Students					0.80
F13	Feeling stimulated to attend e-courses	4.0	0.82	0.51	
F14	Feeling motivated to read provided coursebooks	4.0	0.68	0.59	
F15	Active learning in e-courses	3.9	0.75	0.66	
F16	Adequate personal internet skills	4.1	0.67	0.56	
F17	Adequate personal digital devices	4.0	0.79	0.62	
Output	Satisfied with e-courses in the previous semester				

Note: strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, strongly agree = 5

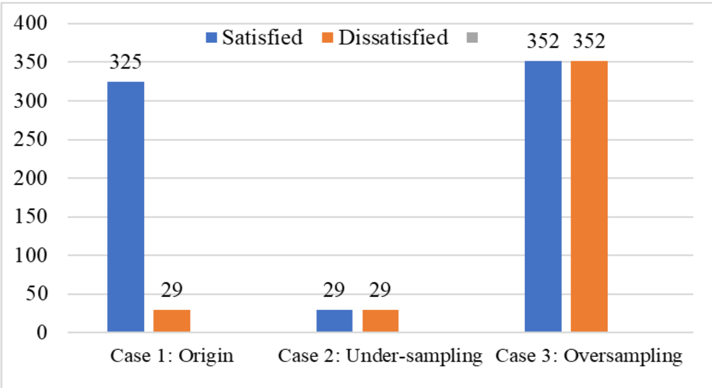


Fig. 2. Data distribution in three classification cases

2.2 Participants

After cleaning data (step 2, section 2.1), valid responses of 499 first-year e-students were selected for further study. Table 2 describes the demographics information of the respondents in this work. Out of 499 students, 344 (68.9%) were women and 155 (31.1%) were men. Regarding details of faculties enrolled, 61 (12.2%) were in the Foreign languages; 48 (9.6%) were in the Mathematics-Informatics Teacher Education; 46 (9.2%) were in the Natural Sciences; 86 (17.2%) were in the Social sciences, Arts, and Humanities; 111 (22.2%) were in the Primary and Pre-school Education; 33 (6.6%) were in the Physical Education - National Security and Defence Education; 85 (17.0%) were in the Economics; 19 (3.8%) were in the Agriculture and environment resource; and 10 (2.0%) were in the Culture Tourism and Social Works. 305 (61.1%) used laptops for e-courses; 22 (4.4%) used desktop computers; 368 (73.7%) used smartphones, and 8 (0.94%) used tablets.

Table 2. Descriptive statistics of the respondents

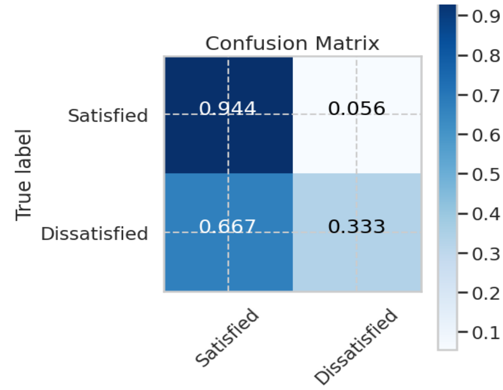
Variable	Number	Percentage
Gender		
Men	155	31.1
Women	344	68.9
Faculties enrolled		
Foreign languages	61	12.2
Mathematics - Informatics teacher education	48	9.6
Natural sciences	46	9.2
Social sciences, arts, and humanities	86	17.2
Primary and pre-school education	111	22.2
Physical education-national security and defense education	33	6.6
Economics	85	17.0
Agriculture and environment resources	19	3.8
Culture tourism and social works	10	2.0
Digital devices used for e-courses		
Laptops	305	61.1
Desktop computers	22	4.4
Smartphones	368	73.7
Tablets	8	0.94

3. RESULTS AND DISCUSSION

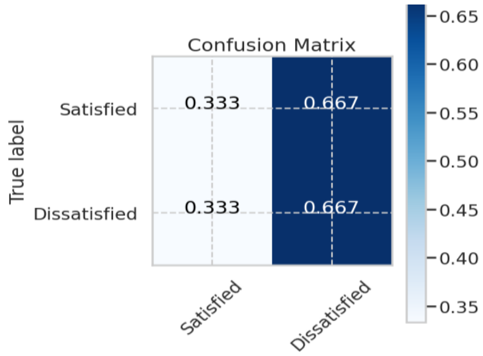
3.1 Classification Results

Table 3 reports the classification results of three data distribution cases. As shown in this table, the results of accuracy, precision, recall, and F1_score of all the constructed ML models in case 3 are higher than those in case 1 and 2. The best model is DT (accuracy = 90.4%; F1_score = 91.2%). LR and SVM gain equal accuracy and F1_score 88.8%. The MLP model achieves an accuracy of 85.0% and F1_score of 83.8%. Thus, DT will be used for further analysis.

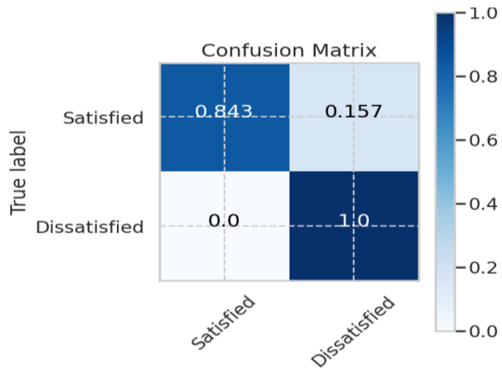
In addition, confusion matrix, ROC curve, and AUC were computed to evaluate the performance of each ML model. Fig. 3 compares the confusion matrix results in three data classification cases. As shown in Fig. 3 (a), this value is 1.0 for "dissatisfied" class and 0.84 for "satisfied" class indicating that the ML models constructed in case 3 are excellent. This means that the DT model built in case 3 can correctly predict dissatisfied and satisfied students. In contrast, this value in case 1 and 2 (Fig. 3 (a), (b)) are low indicating that the DT model built in these two cases cannot classify dissatisfied students correctly.



(a) Case 1: Original



(b) Case 2: Under-sampling



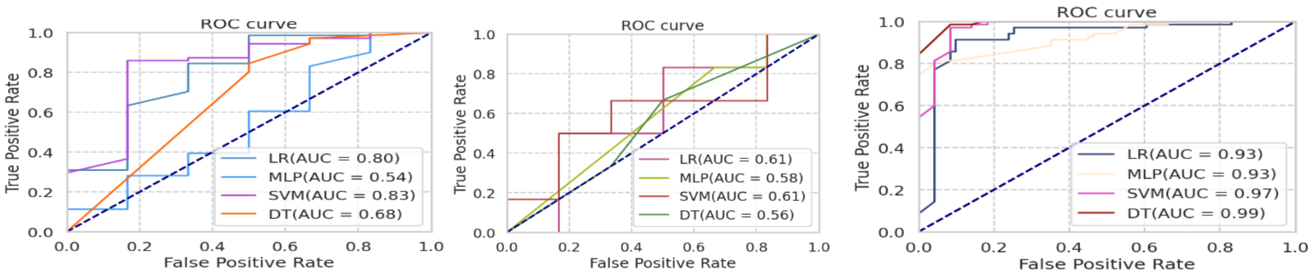
(c) Case 3: Oversampling

Fig. 3. Confusion matrix result of DT (Fold 4)

Fig. 4 benchmarks the results of the ROC-AUC of all four constructed ML models. From Fig. 4 (c), it is clear that in case 3, the ROC-AUC value is close to 1.0 indicating that the ML models constructed in this case are excellent. DT (0.99) is superior to SVM (0.97), MLP (0.93), and LR (0.93). Hence, it is selected to retrieve important features.

Table 3. Classification results

Classification cases	Evaluation matrix	ML models			
		LR %	MLP %	SVM %	DT %
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Case 1: Original	Accuracy	92.0 (0.0)	91.0 (1.0)	92.0 (0.0)	88.0 (2.8)
	Precision	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)
	Recall	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)
	F1_score	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)	00.0 (0.0)
Case 2: Under-sampling	Accuracy	66.6 (16.8)	66.6 (16.8)	80.2 (15.0)	66.6 (19.6)
	Precision	69.4 (20.4)	69.4 (20.4)	84.4 (14.9)	66.0 (19.9)
	Recall	73.4 (18.9)	73.4 (18.9)	73.4 (18.9)	70.0 (24.7)
	F1_score	69.6 (13.6)	69.6 (13.6)	78.2 (16.7)	67.2 (21.4)
Case 3: Oversampling	Accuracy	**88.8 (1.7)**	85.0 (2.1)	**88.8 (1.7)**	**90.4 (2.3)**
	Precision	91.4 (3.3)	90.0 (1.5)	91.4 (3.3)	85.6 (3.8)
	Recall	85.8 (3.7)	78.6 (5.5)	85.8 (3.7)	97.8 (1.6)
	F1_score	**88.8 (2.1)**	83.8 (3.2)	**88.8 (2.1)**	**91.2 (2.0)**



(a) Case 1: Original (b) Case 2: Under-sampling (c) Case 3: Oversampling

Fig. 4. ROC_AUC results of three classification cases

3.2. Discussion

The results showed that the best classification method was oversampling the minority class "dissatisfaction". The outperformed ML model was DT, which gained the highest results of accuracy (90.4%), precision (85.6%), recall (97.8%), F1_score (91.2%), and ROC-AUC (0.9). It was followed by the SVM, LR, and MLP models. The important features were mainly related to technological infrastructure, level of personal computer and Internet skills of students, and teachers' efforts. The outcomes of this work are consistent with the studies of Elliot (2003), Fu (2010), Douglas et al. (2015), Skrbinjek and Dermol (2019), Alnagar (2020), Lu et al. (2020), Sandiwarno et al. (2023), Masadeh et al. (2023).

Fig. 5 displays the rank order of feature importance based on the Gini score in the DT model. For first-year e-students, "easy-access to e-courses in the e-learning platform" (F5) represents the highest important feature; whereas "enthusiastic friendly teacher" (F10) is the lowest impact feature.

Table 4 lists seven top-ranking features, which are closely related to first-year and second-year e-students' dissatisfaction and thus potentially lead to their decisions on dropouts or retention. There is a slight difference between two groups of students.

As shown in Table 4, the seven top-ranking important features associated with first-year e-students' dissatisfaction comprised "easily access e-courses via the school e-learning platform" (F5), "adequate personal internet skills" (F16), "feeling stimulated to attend e-courses (F13), "stable and uninterrupted e-learning platform" (F6), "adequate personal digital devices (F17), "teachers' great efforts to improve students' learning" (F11), and "providing timely responses to students' inquiries" (F12). It is obvious that the interface of e-learning platform, smooth access to the school portal, and uninterrupted e-learning environment are very important for e-freshmen. Students first attend e-courses do not know how to enter the e-courses and who can be contacted to help them out with technical problems. The uneasy e-classes login could lead to demotivation and dropout decisions. In addition, technology-related anxiety and unsure effectiveness of new technology may prevent students from academic achievement (Jon-Chao et al., 2012; Bervell and Umar, 2020). Moreover, for newcomers with inexperience of teaching methods and/or learning activities in e-courses, platform interface, smooth access, and stable environment are the only judging methods for effective e-courses and contribute to possible retention or exit from the

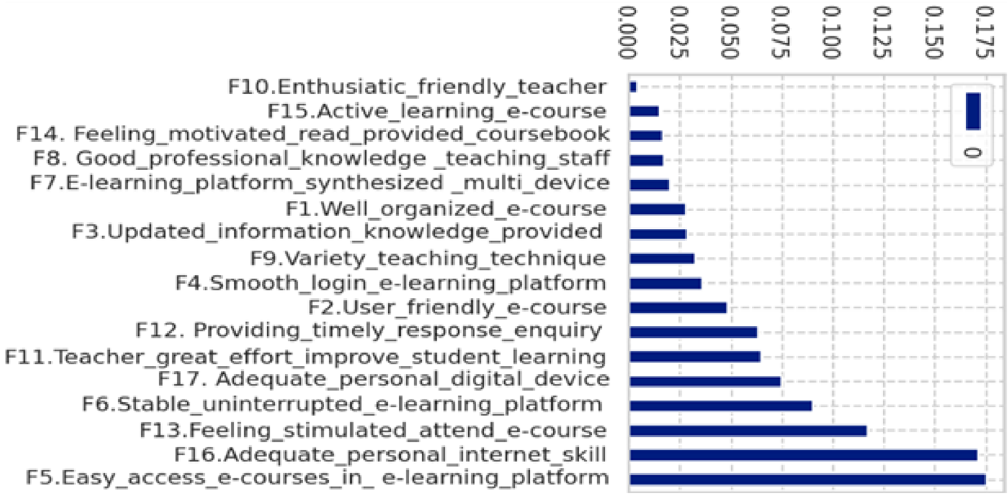


Fig. 5. The rank of feature importance extracted from the DT model

Table 4. The top-ranking important features between first-and second-year e-students

Rank order	First-year e-students	Second-year e-students
1	F5. Easily access e-courses via the school e-learning platform	F17. Adequate personal digital devices
2	F16. Adequate personal internet skills	F16. Adequate personal internet skills
3	F13. Feeling stimulated to attend e-courses	F4. Smoothly login the required e-learning platform
4	F6. Stable and uninterrupted e-learning platform	F9. Variety of teaching techniques
5	F17. Adequate personal digital devices	F13. Feeling stimulated to attend e-courses
6	F11. Teachers' great efforts to improve students' learning	F15. Active learning in e-courses
7	F12. Providing timely responses to student inquiries	F6. Stable and uninterrupted e-learning platform

e-courses / e-learning programs. Complexity of using computers and the operation of e-learning platforms can lead to e-student dissatisfaction and can be a major reason for their withdrawals (Hong, 2002; Puška et al., 2021).

Students, who enter HE, highly expect to complete their studies and achieve a degree diploma. Teaching staff need to make great efforts to better student learning and provide in-time and immediate responses to students' inquiries. Unlike second-year e-students, freshmen first attend e-courses and they may live far away from the campus, the only relation with schools and academic departments is through teachers in e-courses. Teachers' on-time explanations can offer potential to retain first-year e-students.

In order to retain e-students and recruit more e-students, HE should guarantee that the school e-learning platform and environment will be ease of login / access, stable and uninterrupted when offering the courses virtually for substitutable development. Tutorials and 24/7 helpdesk services should be available on the e-learning platforms for easy, comfortable and convenient learning and for information provision. Moreover, the required e-learning platforms should work well in multi digital devices, especially smartphones as more than 73.7% of e-students used smartphones for e-course attendance. Additionally, HE should frequently organize technology training and offer orientation sections to students. Moreover, teachers should use flexible and interesting teaching techniques for enjoyable learning atmospheres. The endless texts and long and/or boring presentations will lead first-year e-students to demotivation and/or exit from the degree completion.

4. CONCLUSION

E-student retention in e-educational programs and e-courses significantly offers the potential for HE to achieve better prestigious and academic quality and finances. It also provides implications for institutional performance nationwide and worldwide. E-student retention, especially first-year e-students, was found to be closely linked to satisfying or dissatisfying experiences in e-educational programs and e-courses. Greater percentages of first-year e-student retention and smaller withdrawals can be gained through ML methods, which are capable of providing HE early warning systems (EWS) for identifying important features associated with dissatisfaction.

The present work successfully constructed an EWS and determined important features associated with first-year e-student dissatisfaction. The results showed that the DT model outperformed SVM, LR, and MLP models. The findings indicated that the most important features associated with e-student dissatisfaction were mainly related to technological infrastructure, students' computer and Internet skill levels as well as teachers' efforts. The remarkable contributions of the present work to e-student dissatisfaction practices and theories can be summarized as follows:

Firstly, this research constructed ML models acting as an EWS, which succeeded in early identification and intervention and helped HE alleviate the dissatisfaction issues for enhancing e-student retention. The constructed ML models allow HE to detect first-year e-students who seem to be dissatisfied before e-students starting to learn for maximizing retention percentage. In contrast, diverse previous related works used assessment, testing, final exams, e-classroom involvement requirements, and students' engagement in e-learning systems features for predictions. These methods failed to offer HE enough time to intervene in first-year e-student dropouts caused by dissatisfaction timely.

Specifically, the present research provided several practical suggestions for minimizing e-student dissatisfaction withdrawals and enhancing e-student recruitment and retention in accordance with the evidential research findings.

Finally, contribution is mainly for theories on predictions on first-year e-student dissatisfaction, which highly leads to potentials of their withdrawals, has currently gained limited attention in theories and practices. It also contributes to the development of effective re-sampling methods to minimize the negative effects of low-accurate dissatisfaction predictions due to imbalanced problems.

Even though this study significantly contributes to practices and theories on first-year e-student dissatisfaction, there were several limitations which lead to potential future works. The present study offers EWS and practical insights within one remote public university in Vietnam, but the adopted methodology can be adapted in various contextualization and with other HE worldwide. The present research only focused on first-year e-students' dissatisfaction by using four factor dimensions: e-course organization, technological infrastructure, teachers, and students and four algorithms to build ML models. Future studies can utilize other features such as assessment, support of non-teaching staff, and secure good career potentials for constructing ML models. Another potential work in the future can be a combination of ML methods and fuzzy logic methods to compare with our outcomes. Efforts to improve the quality of e-student satisfied and dissatisfied experiences as well as retention can be applied to bachelors, masters, and PhDs.

ACKNOWLEDGMENT

This work would like to thank the supports of Chaoyang University of Technology, Taiwan and Dong Thap University, Vietnam. This work was also supported in part by the National Science and Technology Council, Taiwan under Grant NSTC 112-2410-H-324-004.

REFERENCES

Alnagar, D.K.F. 2020. Using artificial neural network to predicted student satisfaction in e-learning. American Journal of Applied Mathematics and Statistics, 8, 90–95.

Amkor, A., Aboulkacem, A., El Bannay, O. El Barbri, N. 2024. Electronic tool coupled with machine learning algorithms for the detection of deltamethrin residues in Mentha Spicata L. International Journal of Applied Science and Engineering, 21(1), 2023265.

Anderson, R.E., Srinivasan, S.S. 2003. E-satisfaction and e-loyalty: A contingency framework. Psychology & Marketing, 20, 123–138.

Baber, H. 2020. Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of COVID-19. Journal of Education and E-learning Research, 7, 285–292.

Bervell, B., Umar, I.N. 2020. Blended learning or face-to-face? Does Tutor anxiety prevent the adoption of Learning Management Systems for distance education in Ghana?. Open Learning: The Journal of Open, Distance and E-learning, 35, 159–177.

Bossman, A., Agyei, S.K. 2022. Technology and instructor dimensions, e-learning satisfaction, and academic performance of distance students in Ghana. Heliyon, 8.

Browne, M.W. 2000. Cross-validation methods. Journal of Mathematical Psychology, 44, 108–132.

Chang, J.-R., Chen, L.-S., Chang, C.-W. 2020. New term weighting methods for classifying textual sentiment data. International Journal of Applied Science and Engineering, 17(3), 257–268.

Chen, L.-S., Su, C.-T. 2008. Using granular computing model to induce scheduling knowledge in dynamic manufacturing environments. International Journal of Computer Integrated Manufacturing, 21(5), 569–583.

Chen, W.-K., Chen, L.-S., Pan, Y.-T. 2021. A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. Applied Soft Computing, 111, 107704.

De Melo Pereira, F.A., Ramos, A.S.M., Gouvêa, M.A., da Costa, M.F. 2015. Satisfaction and continuous use intention of e-learning service in Brazilian public organizations. Computers in Human Behavior, 46, 139–148.

Douglas, J.A., Douglas, A., McClelland, R.J., Davies, J. 2015. Understanding student satisfaction and dissatisfaction: An interpretive study in the UK higher education context. Studies in Higher Education, 40, 329–349.

Elliot, K.M. 2003. Key determinants of student satisfaction. Journal of College Student Retention: Research, Theory and Practice, 4, 271–279.

Fu, F.L. 2010. Comparison of students' satisfaction and dissatisfaction factors in different classroom types in higher education. International Conference on Hybrid Learning, 415–426.

Ho, I.M.K., Cheong, K.Y., Weldon, A. 2021. Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. PLOS ONE, 16, e0249423.

Hong, K.S. 2002. Relationships between students' and instructional variables with satisfaction and learning from a Web-based course. Internet and Higher Education, 5, 267–281.

Huynh-Cam, T.-T., Chen, L.-S., Le, H., 2021. Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance, Algorithms, 14(11), 318.

Huynh-Cam, T.T., Chen, L.S., Ho, V.T. Lu, T.C. 2023. Identifying crucial factors for e-students' satisfaction in Vietnamese emergency remote learning using machine learning methods. In 2023 12th International Conference on Awareness Science and Technology (iCAST), 16–21.

Jameel, A.S., Hamdi, S.S., Karem, M.A., Raewf, M.B. 2021. E-satisfaction based on E-service quality among university students. Journal of Physics: Conference Series, 1804, 012039.

Johnson, J.B., Reddy, P., Chand, R., Naiker, M. 2021. Attitudes and awareness of regional Pacific Island students towards e-learning. International Journal of Educational Technology in Higher Education, 18, 1–20.

Jon-Chao, H., Chan-Jer, H., Chien-Yun, D., Ming-Yueh, H., Pei-Hsin, L., Lee, C.C. 2012. Technology anxiety and implicit learning ability affect technology leadership to promote the use of information technology at elementary schools. Procedia-Social and Behavioral Sciences, 64, 555–563.

Kumar, N., Yadav, S.C. 2023. Comprehensive analysis of deep learning-based text classification models and applications. International Journal of Applied Science and Engineering, 20(3), 2022342.

Lu, D.N., Le, H.Q., Vu, T.H., 2020. The factors affecting acceptance of e-learning: A machine learning algorithm approach. Education Sciences, 10, 270.

Martín Rodríguez, Ó., González-Gómez, F., Guardiola, J. 2019. Do course evaluation systems have an influence on e-learning student satisfaction? Higher Education Evaluation and Development, 13, 18–32.

Masadeh, R.E., Almajali, D., Alrowwad, A.A., Alkhawaldeh, R., Khwaldeh, S., Obeid, B., 2023. Evaluation of factors affecting university students' satisfaction with e-learning systems used during Covid-19 crisis: A field study in Jordanian higher education institutions. International Journal of Data and Network Science, 7, 199–214.

Muljana, P.S., Luo, T., 2019. Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. Journal of Information Technology Education: Research, 18, 19–57.

Nevill, A., Rhodes, C. 2004. Academic and social integration in higher education: a survey of satisfaction and dissatisfaction within a first-year education studies cohort at a new university. Journal of Further and Higher Education, 28, 179–193.

Pérez-Pérez, M., Serrano-Bedia, A.M., García-Piqueres, G. 2020. An analysis of factors affecting students' perceptions of learning outcomes with Moodle. Journal of Further and Higher Education, 44, 1114–1129.

Pham, L., Limbu, Y.B., Bui, T.K., Nguyen, H.T., Pham, H.T. 2019. Does e-learning service quality influence e-learning student satisfaction and loyalty? Evidence from Vietnam. International Journal of Educational Technology in Higher Education, 16, 1–26.

Pham, P.T., Duong, T.B., Phan, T.T.T., Nguyen, T.H., Nguyen, M.T., Tuyet, T.L.T., Hoang, N.D., Yen, D.H., Nguyen, T.T. 2020. Managing continuing education via distance learning and face-to-face courses for human resource development in the Mekong Delta of Vietnam. International Journal of Learning, Teaching and Educational Research, 19, 150–171.

Puška, A., Puška, E., Dragić, L., Maksimović, A., Osmanović, N. 2021. Students' satisfaction with e-learning platforms in Bosnia and Herzegovina. Technology, Knowledge and Learning, 26, 173–191.

Sandiwarno, S., Niu, Z., Nyamawe, A.S. 2023. A novel hybrid machine learning model for analyzing e-learning users' satisfaction. International Journal of Human-Computer Interaction, 1–22.

Schulz, M., 2023. E-learning as a development Tool. Sustainability, 15, 15012.

Skrbinjek, V., Dermol, V., 2019. Predicting students' satisfaction using a decision tree. Tertiary Education and Management, 25, 101–113.

Sung, P.C. 2022. Efficacy of real-time audio biofeedback on physiological strains for simulated tasks with medium and heavy loads. International Journal of Applied Science and Engineering, 19(4), 2022308.

Ullah, M.A., Alam, M.M., Mahiuddin, M., Rahman, M.M. 2019. Predicting factors of students dissatisfaction for retention. Emerging Technologies in Data Mining and Information Security, Springer, Singapore.

Violante, M.G., Vezzetti, E. 2015. Virtual interactive e-learning application: An evaluation of the student satisfaction. Computer Applications in Engineering Education, 23, 72–91.

Wang, X, Hassan, A.B., Pyng, H.S., Ye, H. 2023. Development and empirical study of international student satisfaction model of online course learning interaction in Chinese universities. Education and Information

International Journal of Applied Science and Engineering

Technologies, 28, 16951–16977.

Yang, Z., Becerik-Gerber, B., Mino, L. 2013. A study on student perceptions of higher education classrooms: Impact of classroom attributes on student satisfaction and performance. Building and Environment, 70, 171–188.

Yoke, L.B. 2018. Interrelationship between perceived instructor power, student dissatisfaction, and complaint behaviors in the context of higher education. International Education Studies, 11, 12–25.