Received 16 August 2021
Revised 5 May 2022
Accepted 5 May 2022

Evaluating sustainability of mobile learning framework for higher education: a machine learning approach

Dhammika Manjula Dolawattha
University of Kelaniya, Kelaniya, Sri Lanka
H.K. Salinda Premadasa
Centre for Computer Studies, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka, and
Prasad M. Jayaweera
University of Sri Jayewardenepura, Nugegoda, Sri Lanka

Abstract

Purpose – The purpose of this study is to evaluate the sustainability of the proposed mobile learning framework for higher education. Most sustainability evaluation studies use quantitative and qualitative methods with statistical approaches. Sometimes, in previous studies, machine learning models were utilized conventionally.

Design/methodology/approach – In the proposed method, the authors use a novel machine learning-based ensemble approach with severity indexes to evaluate the sustainability of the proposed mobile learning system. In this severity indexes, consider the cause-and-effect relationship to identify the hidden correlation among sustainability factors. Also, the proposed novel sustainability evaluation algorithm helps to evaluate and improve sustainability iteratively to have an optimal sustainable mobile learning system. In total, 150 learners and 150 teachers in the university community engaged in the study by taking the sustainability questionnaire. The questionnaire consists of 20 questions that represent 20 sustainable factors in five sustainability dimensions, i.e. economic, social, political, technological and pedagogical.

Findings – The results reveal that the proposed system has achieved its economic and pedagogical sustainability. However, the results further reveal that the proposed system needs to be improved on technological, social and political sustainability.

Originality/value – The study focused novel machine learning approach and technique for evaluating sustainability of the proposed mobile learning framework.

Keywords Sustainable mobile learning, Machine learning, Artificial neural network, Support vector machine, Ensemble classifier, Sensitivity analysis

Paper type Research paper

1. Introduction

Sustainable learning systems capable to provide learning services to their stakeholders long lastly in diverse conditions (Hays and Reinders, 2020). The American Society of Civil Engineers (ASCE) defines sustainability as "a set of economic, environmental and social conditions in which all of society has the capacity and opportunity to maintain and improve its quality of life indefinitely without degrading the quantity, quality or the availability of economic, environmental and social resources. Sustainable development is the application of these resources to enhance the safety, welfare, and quality of life for all of society" (ASCE, 2018). A learning system can be considered as sustainable if it is satisfied with human sustainability: Individual needs should be protected and supported with dignity and in a way that developments should improve the quality of human life and not threaten human beings; social sustainability: Relationships of people within a society should be equitable, diverse,

The International Journal of Information and Learning Technology
Vol. 39 No. 3, 2022
pp. 266-281

2056-4880
DOI 10.1108/IJILT-08-2021-0121

connected and democratic; technical sustainability: Technology must cope with changes and evolution fairly, respecting natural resources; environmental sustainability: Natural resources have to be protected from human needs and wastes; and economic sustainability: A positive economic value and capital should be ensured and preserved (Alharthi *et al.*, 2018). Sustainable learning systems are more efficient and productive than the learning system with a low level of sustainability. Therefore, evaluating sustainability is more useful for the novel learning system. With that motivation, it is essential to evaluate the sustainability of the proposed mobile learning (ML) framework for higher education. In previous studies, sustainability was evaluated for many research areas, i.e. education (Ofei-Manu and Didham, 2018) (Mabila *et al.*, 2017), technology (Coskun-Setirek and Tanrikulu, 2019), agriculture (Luan *et al.*, 2018), production (Venugopal and Saleeshya, 2019), information and communication technology (ICT) (Ng and Nicholas, 2013; Ziemba, 2017), etc. However, in these studies, researchers mostly use quantitative statistical techniques or qualitative approaches to measure sustainability. Besides, researchers use machine learning models to predict the overall sustainability of the system or product (Abdella *et al.*, 2020; Nilashi *et al.*, 2019; Nosratabadi *et al.*, 2019). But, the authors use a novel technique to evaluate the sustainability of the ML framework using an approach in statistical, and ensemble method in machine learning. Also, an algorithm that evaluates the overall sustainability of the framework iteratively was proposed. For that, the authors use the severity index to identify important sustainability factors that require concentrating on enhancing the overall sustainability of the system. Also, the authors hope to use the ML system that is develop by implementing the proposed ML framework for higher education.

The paper is structured as follows: Section 2 describes the related works, Section 3 describes the preliminaries. It includes the details of the proposed ML system, machine learning models with ensemble classifier, which is used in this study, and model evaluation and validation. Section 4 describes the proposed method, and Section 5 describes the methodology of this study. Section 6 shows the results and discussion, and finally, Section 7 describes the conclusion and implications.

2. Related works

Sustainability evaluation of various research areas can be found in previous studies. The researchers apply diverse approaches, tools and techniques to measure sustainability. A framework for ML sustainability was proposed, and it revealed that the dimensions needed for ICT sustainability in education are economic sustainability, social sustainability, political sustainability, technological sustainability and pedagogical sustainability. The implementation was done through ML programs in an educational institute. Statistical and theme-based qualitative methods were utilized to evaluate the study that was done with the participation of 57 students and 25 teachers. The results further revealed that positive attitudes, better communication and trust need to be developed among key players (Ng and Nicholas, 2013). Coskun-Setirek and Tanrikulu (2019) explore the technological sustainability of ML using 11 ML admins and 75 ML staff members with a statistical data analysis approach. They found that quality standards, requirement specification, expansion and upgrade, and maintenance issues are required to be improved, for technological sustainablity of ML while accessibility, interoperability, connectivity and availability of system use support issues have already reached the satisfactory level (Coskun-Setirek and Tanrikulu, 2019). Ziemba (2017) conducted a study to build a theoretical model for ICT adoption to the initiatives in the sustainable information society, and they illustrated that sustainable information society is constructed with factors in ecological, economic, sociocultural and political. In this study, 396 initiatives were analyzed quantitatively, and ICT's quality, management and information culture acted as major roles in the sustainability of a society (Ziemba, 2017). Sensitivity analysis of the indices-based approach is used to evaluate the sustainable development level in agriculture. In this study, a quantitative evaluation system was used, and it has four dimensions, i.e. agricultural economic development level, agricultural productive factors development level, agricultural social development level and agricultural resources and environment development level (Luan *et al.*, 2018). Another framework has been develped for appraising the sustainability in the manufacturing of the Ayurveda pharmaceutical industry. It reduces wastes and reacts to dynamic changes by applying manufacturing strategies such as lean and agile through the sustainability dimensions such as economic, environmental, social, technological and ethical (Venugopal and Saleeshya, 2019). The United State Agency for International Development (United State Agency for International Development (USAID) funded Girls Improved Learning Outcomes (GILO) program for ICT in education in Egypt was evaluated using a framework of four dimensions of ICT sustainability, i.e. technological, individual and social, economic and political. The evaluation process included the use of interviews, document reviews and follow-up phone calls with the schools and school staff (Pouezevara *et al.*, 2014). Machine learning-based sustainability evaluating techniques can be found in previous studies. Abdella *et al.* (2020) carried out a machine learning-based study to find sustainability in the food industry. In this study, K-mean clustering and logistic regression-based model are used to assess and model the various dimension of sustainability in food consumption such as environment, economic and social (Abdella *et al.*, 2020). A machine learning-based sustainability assessment method is used to predict the overall sustainability of a country. The fuzzy clustering-centered machine learning approach is attempted to identify the correlation among human, ecological and overall sustainability performances of a nation. This method utilizes sustainability details of 128 countries to forecast the sustainability of a country using any number of sustainability indicators (Nilashi *et al.*, 2019). A study was carried out to explore the current usage of machine learning and deep learning methods in the sustainability of smart cities. The survey reveals that power, well-being and urban carriage are useful areas that render the models (Nosratabadi *et al.*, 2019). The main drawbacks of existing sustainability evaluating methods are as follows. They rely on checklist- or questionnaire-type sustainability measures by calculating the mean scores or measure the sustainability through ideas of experts or end-users. They do not provide any numerical evidence by ranking the importance of sustainability factors. Here, the authors proposed not only the sustainability evaluation method but also proposed which factors should be considered first to improve the sustainability of the ML system. For that, investigations are carried to find the order of importance of sustainability factors through sensitivity measures in the machine learning models.

3. Preliminaries

In this section, key techniques and tools behind the proposed method are discussed. The ML framework, which will be evaluated for sustainability in this study, is discussed first. Next, which machine learning models are used as prediction models in this study is discussed. The next method of evaluating machine learning models is discussed to find the optimal prediction model. Finally, the prediction model validating technique is elaborated.

3.1 Mobile learning framework for higher education

In this study, the authors intend to evaluate the sustainability of the ML system implemented based on the proposed ML framework for higher education (MLFrame). The main modules (components or independent variables) of this framework are learner, teacher, ML devices, ML tools, ML contents, higher education institutes and communication technology (Figure 1). These modules consist of several influencing factors such as motivation, usefulness, interactivity, ease of use, etc. Different ML facilities are integrated such as chat, forum, games, quizzes, assignments, etc. to the MLFrame are realized above influencing factors. MLFrame is implemented through the Moodle ML environment (Moodle, 2020). For that, new features are integrated into Moodle mobile application by enhancing the existing Moodle plugins (Dolawattha *et al.*, 2019). Hence, Moodle mobile application is customized to implement the facilities introduced in the MLFrame by enabling the existing Moodle plugins in Moodle desktop version to serve academic functionalities in the Moodle mobile environment.

3.2 Machine learning models

In this study, few supervised machine learning algorithms are used to build predictor models and find hidden relationships between input and output variables.

An artificial neural network (ANN) consists of computational algorithms and mimics the human brain transactions. It comprises connected individual processing units calls neurons in three layers, i.e. input, hidden and output. The input layer is used to feed the values of input variables to be modeled. ANN consists of one or more hidden layers. Each hidden layer consists of neurons with values computed by using individual connection weight and input values. The output layer consists of one or more output nodes that correspond to the prediction to be done. Output nodes are connected to input nodes via hidden layer nodes with combination functions and transfer functions. Also, activation functions are used to pass values from one neuron to another (Keller *et al.*, 2016).

Support vector machine (SVM) classifies data points distinctly by discovering hyperplane in N-dimensional space (for N number of input variables or features) (Pereira and Borysov, 2019).

Decision tree (DT) – The DT has a tree-like structure; each root node denotes a condition for feature (or input variable), each leaf node denotes class label (or final decision) and each branch denotes roads to final decision and makes class label by satisfying each root node condition (Vaughn, 2018).

3.3 Majority voting ensemble classifier

An ensemble classifier is a collection of classifiers whose separate outputs are joined to obtain better output than the output of a considered collection of classifiers. These classifiers are joined by weighted or unweighted voting techniques (Dietterich, 2000). The concept behind using a classifier as an ensemble method is to make a prediction based on majority voting. If authors have a training data set, n classifiers or classification models, and each classifier was trained on the training set (Raschka and Mirjalili, 2017), Figure 2 shows the working process of the majority voting ensemble classifiers.

Initially, the training data set used to train the classifiers then, it can be made the prediction using trained classifiers. Classifiers h_1, h_2, …. h_n make prediction $\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n$. Each classifier gives one prediction for a new data. Therefore n-number of predictions. Then the voting scheme, which calls the majority voting uses to decide the final prediction

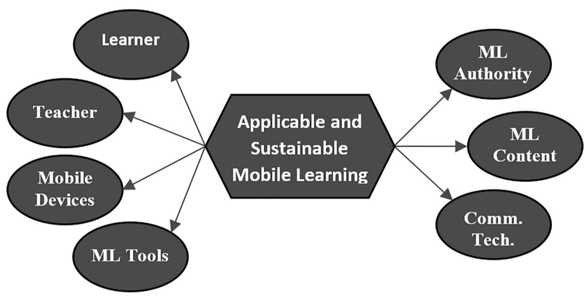

Figure 1.
The ML framework for higher education (MLFrame)

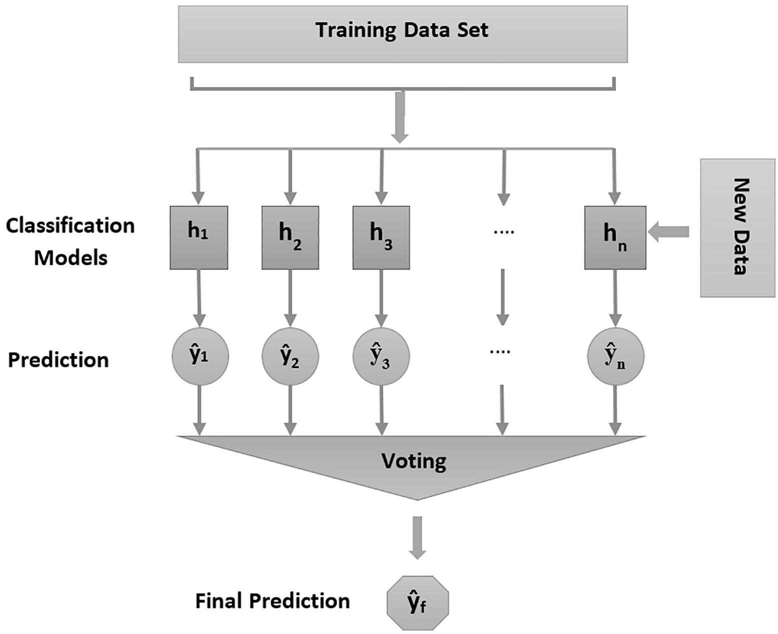

Figure 2.
Majority voting ensemble method

(Figure 2). Voting is reducing end class label predictions for the single data point into a single class label. So, mode is applied to get the final class label.

$$\hat{y}_f = \text{mode}\{h_1(x), h_2(x), \ldots h_n(x)\}, \text{ where } h_i(x) = \hat{y}_i \quad (1)$$

3.3.1 Importance of majority voting. Assume, if we have;

(1) n independent classifiers $h_1, h_2, \ldots h_n$ with base error rate $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, and they are uncorrelated;

(2) A binary classification task (i.e. has only two class labels, class label1 and class label2); and

(3) The error rate is better than random guessing (i.e. lower than 0.5 for binary classification).

$$\forall \varepsilon_i \in \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}, \varepsilon_i < 0.5$$

The probability that makes a wrong prediction via the ensemble, if k classifiers predict the same class label, according to the probability mass function of a binomial distribution is:

$$P(k) = \binom{n}{k} \varepsilon^k (1-\varepsilon)^{n-k} \quad k > n\Big/2 \quad (2)$$

Ensemble error, according to the cumulative probability distribution, is:

$$\varepsilon_{ens} = \sum_{k}^{n} \binom{n}{k} \varepsilon^k (1-\varepsilon)^{n-k} \quad (3)$$

When base error ε $(0 < \varepsilon < 0.5)$, the ensemble error < base error.

Therefore, ensemble error is always less than the error rate of a single classifier (Figure 3). Finally, the majority voting ensemble classifier has better performance than a single classifier (Kang *et al.*, 2017).

3.4 Model evaluation

Model (or classifier) evaluation is a continuous process and very important when developing a prediction model because it certifies the model's best fit and future performance for data. In this study, the authors choose popular evaluation techniques such as mean squared error (MSE) and correlation.

MSE – this is utmost far and wide used and operative performance function. It can be defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - F_i)^2 \qquad (4)$$

where $Y_i =$ The actual observation for the target variable.

$F_i =$ The predicted value by the model.

$n =$ Total number of data points in the testing data set

Correlation – correlation of testing data set can be denoted as:

$$r_{F_i,\ Y_i} = \sum_{i=1}^{n}\frac{\left(F_i - \overline{F}_i\right)\left(Y_i - \overline{Y}_i\right)}{(n-1)s_{F_i} \times s_{Y_i}} \qquad (5)$$

where:

$\overline{F}_i =$ mean of predicted values;

$\overline{Y}_i =$ mean of actual observations;

$s_{F_i} =$ standard deviation of predicted values; and

$s_{Y_i} =$ standard deviation of actual observation.

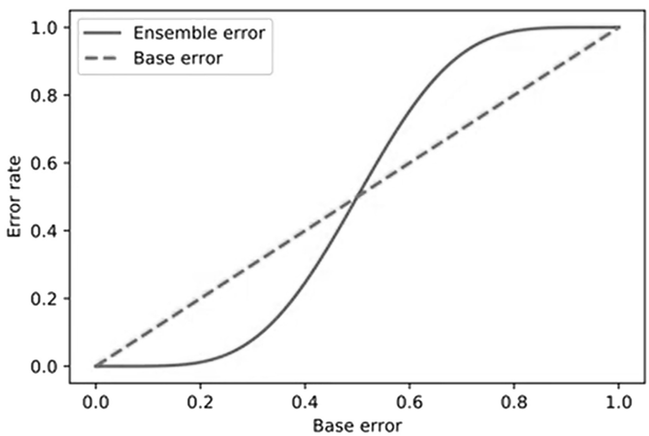

Figure 3.
Error rate of ensemble classifier with a base classifier (Raschka and Mirjalili, 2017)

Here, the correlation between actual user response and model-predicted value for the output variable is considered. In this kind of human-related studies, the correlation is required to be above 0.3 for better prediction (Cohen *et al.*, 2013) and the larger correlation valued model is the best performance model (Makridakis *et al.*, 1998), while the smaller the MSE is considered as the better prediction model.

3.5 Model validation

This is the method for minimizing the bias associated with the random sampling of the training and holdout data samples in comparing the predictive accuracy.

Rotation estimation – the total data set (D) is arbitrarily fragmented into k mutually exclusive subset (the folds: $D_1, D_2, \ldots, D_k$) of nearly equal size. The model (or classifier) is trained and tested k times; each time ($t \in \{1, 2, \ldots, k\}$), it is trained on all, except one fold (D_t), and tested on the remaining single fold (D_t). The cross-validation estimate of the overall performance criteria is calculated as simply the average of the k individual performance measures as follows:

$$CV = \frac{1}{k}\sum_{i=1}^{k} PM_i \tag{6}$$

where CV = cross-validation, k = number of folds, PM = the performance measure for each fold (Jung, 2018).

4. Proposed method

In this study, the proposed sustainability evaluation algorithm (Figure 4) was used to evaluate the sustainability of the ML system. The authors developed checklist questions/items to get the user responses to measure the overall sustainability of the system. This checklist consists of 20 sustainability factors under five different sustainability dimensions, i.e. economic sustainability, social sustainability, political sustainability, technological

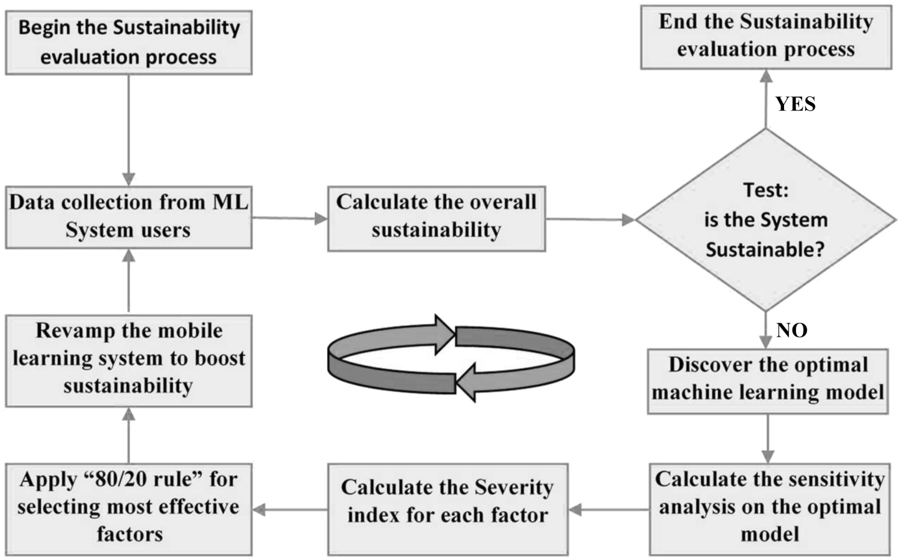

Figure 4.
Proposed sustainability evaluation algorithm

sustainability and pedagogical sustainability. The mean values for each sustainability factor are tested for whether it reaches a satisfactory level for sustainability. This checklist consists of five-point Likert scale with the values, 1 for strongly dissatisfied, 2 for dissatisfied, 3 for normal, 4 for satisfied and 5 for strongly satisfied. For a sustainable ML system, the system should be evaluated with an overall mean value equal to 4 or higher. In conventional evaluation, to improve the overall quality, sustainability processes consider lower evaluated individual factors in checklist questions or items. Even though lower-valued sustainable factors are improved, they cannot be assured of the overall sustainability of the system with the above improvements. Therefore, time and efforts should be attained on notable problems that improve overall system sustainability in the end. As a solution for that, the authors proposed to use the severity index in the sustainability evaluation process, which considers the hidden importance of sustainable factors through a machine learning model.

$$\text{Severity index} = \text{Sensitivity score} \times \frac{1}{\text{Average of checklist evaluation scores}} \tag{7}$$

where sensitivity score is the importance of the checklist item in predicting sustainability, and it is discussed under Section 4.1. The reciprocal of the average of checklist evaluation scores is used to have a bigger value for fewer values of the average of checklist evaluation scores (mean values of individual sustainability factors) for each sustainability factor. Therefore, if a particular checklist item has a larger sensitivity score with a smaller average of checklist evaluation score, it gives more importance when predicting the overall sustainability of the system.

Overall sustainability is included as the last item in the sustainability checklist questions/items and used as the output variable, while the other 20 checklist questions/items are used as input variables to developed prediction models. The first step in this methodology is gathering responses for the checklist from the sample of target users (Figure 4). Then, calculate the overall sustainability when the overall sustainability is greater than or equal to Likert scale value 4, it is concluded that the system is sustainable, it is concluded that the system is sustainable. Otherwise, further steps are required to continue as in the proposed sustainability evaluation algorithm. So, the next step is investigating the best predictive model that explains the hidden complex relationship between input variables and the output variable. For that, various performance measures are taken into account in machine learning models (or ensemble classifier) such as MSE and correlation. Using this identified best predictive model (or classifier), a sensitivity analysis is carried out for input variables (checklist items). According to equations (8) and (9), to calculate the sensitivity score, MSEs of the selected prediction model for each sustainability factor were determined. For that mean squared errors of the selected prediction model for each sustainability factor were calculated with the absence of the considered sustainability factor each time. Then, checklist items are ranked in descending order by developing the severity index as mentioned in equation (7). This ranking will be used to realize which variable will be improved first with limited resources. Finally, apply the 80/20 rule for selecting the most effective factors. These selected effective factors address nearly 80% of sustainability problems in the ML framework from 20% of causes (Harvey, 2018). The proposed sustainability evaluating algorithm is an iterative process that needs to continue until the system obtains a satisfying level of sustainability.

4.1 Sensitivity score

The best prediction model (or classifier) can be selected using performance criteria such as MSE and correlation. Then, the ranking order for the importance of each independent variable can be identified by using the above-selected prediction model (or classifier). The sensitivity score of each predictor variable is determined by doing sensitivity analysis. Sensitivity analysis is the method of identifying cause-and-effect relationships among the inputs and outputs. The main idea behind the sensitivity analysis is that the performance change in the predictor model with the absence of the particular predictor variable will be conducted the sensitivity analysis (Principe *et al.*, 2001; Principe *et al.*, 2000). Hence, the measure of sensitivity or sensitivity score of a particular predictor variable is the ratio of the error of the model (or classifier) with the absence of the predictor variable to the error of the model (or classifier) that contains the predictor variable. Sensitivity measure of *i*th predictor variable (Saltelli, 2002) is:

$$S_i = \frac{V_i}{V(F_t)} = \frac{V(E(F_t X_i))}{V(F_t)} \tag{8}$$

where $V(F_t)$ = unconditional output variance, $V(E(F_t X_i)$ = overall output variance but X_i

Therefore, the sensitivity measure of *i*th predictor variable can be written by using equations (4) and (8).

$$S_i = \frac{MSE(E(X_i))}{MSE} \tag{9}$$

where MSE = mean squared error, $MSE(E(Xi))$ = mean squared error but X_i (mean squared error of the model without X_i).

4.2 Sustainability dimension and factors

By literature, five dimensions and 20 sub-factors for evaluating sustainability of the system were found (Table 1). According to them, the authors proposed sustainability questions/items for evaluating the system with the participation of end-users (Table 2).

5. Methodology

The ML system developed based on the proposed ML framework for higher education was used in this study. Initially, 150 university students and 150 university teachers belonging to five different faculties of a state university, i.e. faculties of Science, commerce and management, social sciences, humanities and medicine, were selected as the sample. First of all, intended users were asked to use the ML system through their mobile devices. Also, users were asked to use the app for their academic activities such as developing online courses,

Dimension	Factors
Technological (Pouezevara *et al.*, 2014), (Ng and Nicholas, 2013), (Mabila *et al.*, 2017), (Coskun-Setirek and Tanrikulu, 2019)	Maintenance, assistance, newest technology, adjustability
Social (Pouezevara *et al.*, 2014), (Ofei-Manu and Didham, 2018), (Ng and Nicholas, 2013), (Mabila *et al.*, 2017), (Ziemba, 2017)	Useful to join, collaborative, awareness, enough technology availability
Economic (Pouezevara *et al.*, 2014), (Ofei-Manu and Didham, 2018), (Ng and Nicholas, 2013), (Mabila *et al.*, 2017), (Ziemba, 2017)	Cost for device, cost for connectivity, cost for educational environment, cost for maintenance, institutional infrastructure cost
Political (Pouezevara *et al.*, 2014), (Ofei-Manu and Didham, 2018), (Ng and Nicholas, 2013), (Ziemba, 2017)	Supports of university management, align with national educational policy, supports of ministry, supports of university grant commission
Pedagogical (Ng and Nicholas, 2013), (Mabila *et al.*, 2017)	Instructional methodology, interactive content, efficiency, effectiveness

Table 1. Sustainability dimensions and factors for evaluating the sustainability of the proposed system

Dimensions, abbreviation and factors		Corresponding sustainability question/item
Economic sustainability		
ECS1-DC	Device cost	Device cost is reasonable
ECS2-CC	Connectivity cost	Connection charges bearable
ECS3-SC	Software cost	Support software is free or low-cost
ECS4-IC	Institution capability	The institution can manage expenses for all the educational services
Social sustainability		
SOS1-CO	Collaboration	Provide necessary facilities for collaborative learning with the peers
SOS2-SH	Sharing	Provide necessary facilities for information sharing for developing new knowledge
SOS3-IN	Influence	The influence of the institutional community is high for use of the ML system
SOS4-AC	Acceptance	Acceptance of the institutional community is high for use of the ML system
Political sustainability		
POS1-PE	Political environment	It has a suitable political background to pursue academic activities through the ML system
POS2-LE	Leadership	The Head of the institution/system in charge manages background services to pursue academic activities through the ML system
POS3-IP	Institution policy	The institutional policy supports the pursuit of academic activities through the ML system
POS4-IB	Institutional barriers	No barriers within the institution such as teachers, admins reluctant to use technology for learning and teaching
Technological sustainability		
TES1-CO	Connectivity	Network bandwidth is capable enough to pursue academic activities through the ML system
TES2-DE	Device	Device owned supports to pursue academic activities through the ML system
TES3-SU	Support	Support staffs (i.e. academic and technical) provide better services to pursue academic activities through the ML system
TES4-SE	Security	Content, user profile and system are secured
Pedagogical sustainability		
PES1-IM	Instructional methodology	Learning content arrange in the system is appropriate
PES2-IC	Interactive content	Learning content and tools integrated into the system is interactive
PES3-EF	Efficiency	Teaching/learning efficiency in the system is optimal
PES4-EF	Effectiveness	Teaching/learning effectiveness in the system is satisfied

Table 2. Sustainability questions/items with their shortened symbols

assignment submission, assignment grading, forum discussion, group discussion via chat facility, etc. Users accessed the mobile application around 1.5 months before they were asked to respond to the sustainability checklist questions/items. Finally, 300 user responses were analyzed. In this analysis, pre-processing activities were done to complete missing responses and rectify wrong responses done mistakenly. The average mean score for each sustainability factor was calculated. Ten-fold cross-validation was used to train and test each prediction model. MSE and correlations were calculated for each of the three models. Python programming language with scikit-learn free machine learning code library is used to implement the machine learning algorithms and ensemble classifier used in this study. Sensitivity analysis was done, and the sensitivity score was calculated for each predictor variable using the best-performed prediction model (or ensemble classifier). That means, 20 different MSEs of the best prediction model were calculated with the absence of each predictor (sustainability) factor (as described in the proposed method in Section 4). Then, the severity index for each predictor variable was calculated. Finally, the "80/20" rule was applied for identifying the most effective sustainability factors.

6. Results and discussion

Average checklist item scores are mentioned in Figure 5, and the overall mean value of sustainability factors is 3.88725 (mean value of average checklist score using the second column in Table 4). That means the system is not sustainable and needs to do sustainability improvements following the sustainability algorithm mentioned in Figure 4. According to the graph of the average checklist item score, the conventional process for improving the sustainability of the learning system should be improved by the sequence of sustainability factors ECS2-CC, TES4-SE, TES3-SU.

In this study, the proposed method is used to evaluate sustainability by calculating the severity index of each sustainability factor using a machine learning algorithm. For that, questionnaire responses were used to train machine learning algorithms such as ANNs, SVMs and DTs. In the questionnaire, 20 different sustainable factors were used as predictor variables or input variables. Also, user-evaluated overall sustainability is used as an output variable in the training process in each machine learning model. The ten-fold cross-validation method is used 1,000 times iteratively to have optimal training for machine learning models. The best predictive model was selected by considering model evaluation using MSE and correlation. The model evaluation results were mentioned in Table 3. Among the three base prediction models, the ANN performs the best correlation between input variables and output variables. But, their ensemble classifier performs the greatest correlation than any single base classifier.

According to the criteria specified in Section 3.3 and the results of Table 3, the ensemble classifier model is the better prediction model for investigating the sustainability of the ML system by studying hidden relationships that lie among sustainability factors. The ensemble classifier model trained by using the user responses of the sustainability evaluation checklist

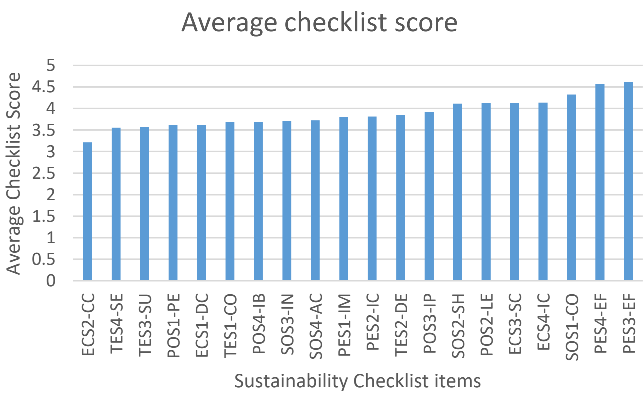

Figure 5.
Graph of average checklist scores

was used to do the sensitivity analysis for prediction variables. However, among the three base classifiers, the neural network shows the best results. Figure 6 shows a typical illustration of a multilayer perceptron-type neural network used in this study. It includes 20 neurons in the input layer, 14 and 7 neurons in the first and second hidden layers, respectively. Only one output variable consists of the output layer for the overall sustainability of the system.

The sensitivity score (or measure of sensitivity score) and sensitivity index were calculated using the ensemble classifier. Table 4 shows the results of sensitivity analysis and severity index calculation. The rule "80/20" was applied to select the vital sustainability items to improve the overall sustainability of the ML system, which was implemented based on the proposed applicable and sustainable ML system for higher education.

Figure 7 depicts the Pareto chart and presents the severity of each sustainability item in descending order. This is the order of importance for improving the sustainability of the system suggested by the ensemble classifier as the best machine learning prediction model. However, a significant difference shows for this order of importance between our proposed machine learning techniques and the conventional average checklist scores mentioned in Figure 5.

The original "80/20" rule suggested considering checklist items that cover 80 cumulative percentages for improving the sustainability of system (15 items starting from sustainability checklist items from TES2-DE to ECS3-CS in Figure 7). Here, the authors propose to select the

	Performance measures	
Prediction model	MSE	Correlation
ANN	0.084	0.771
SVM	0.122	0.514
DTs	0.092	0.741
Ensemble classifier	0.068	0.786

Table 3. Ten-fold cross-validation model results

Input variable	Average checklist score	MSE(E(X_i))	Measure of sensitivity	Severity index
ECS1-DC	3.621	0.061	0.897058824	0.247737869
ECS2-CC	3.212	0.063	0.926470588	0.288440407
ECS3-SC	4.124	0.081	1.191176471	0.288840075
ECS4-IC	4.134	0.071	1.044117647	0.252568371
SOS1-CO	4.321	0.101	1.485294118	0.343738514
SOS2-SH	4.112	0.084	1.235294118	0.300411994
SOS3-IN	3.712	0.111	1.632352941	0.439750254
SOS4-AC	3.723	0.074	1.088235294	0.292300643
POS1-PE	3.612	0.081	1.191176471	0.329783076
POS2-LE	4.123	0.121	1.779411765	0.431581801
POS3-IP	3.911	0.115	1.691176471	0.432415359
POS4-IB	3.689	0.079	1.161764706	0.314926730
TES1-CO	3.682	0.089	1.308823529	0.355465380
TES2-DE	3.855	0.121	1.779411765	0.461585412
TES3-SU	3.567	0.111	1.632352941	0.457626280
TES4-SE	3.552	0.091	1.338235294	0.376755432
PES1-IM	3.805	0.095	1.397058824	0.367163948
PES2-IC	3.811	0.092	1.352941176	0.355009493
PES3-EF	4.612	0.072	1.058823529	0.229580123
PES4-EF	4.567	0.074	1.088235294	0.238282307

Table 4. Severity index calculations

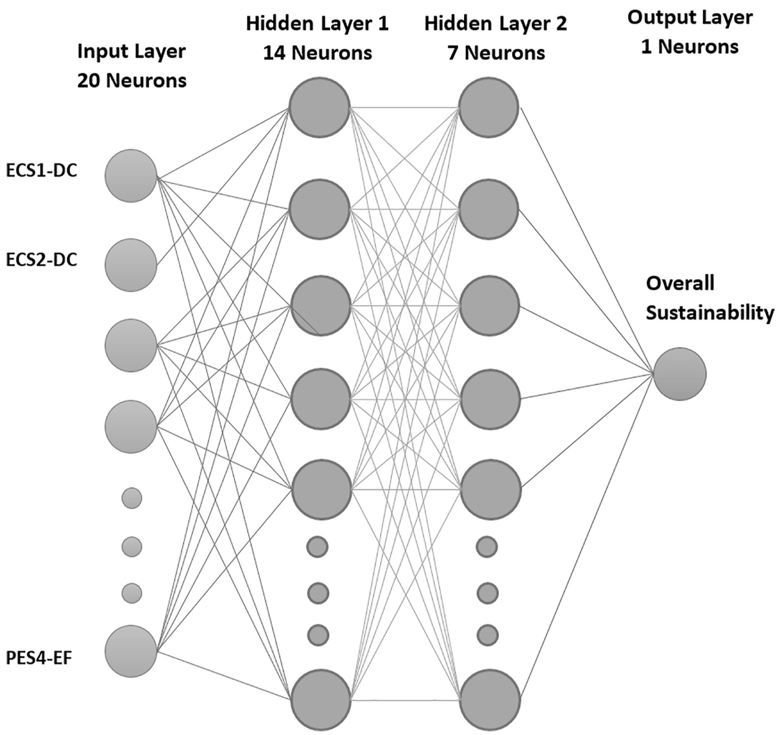

Figure 6.
ANN architecture used in the ensemble classifier

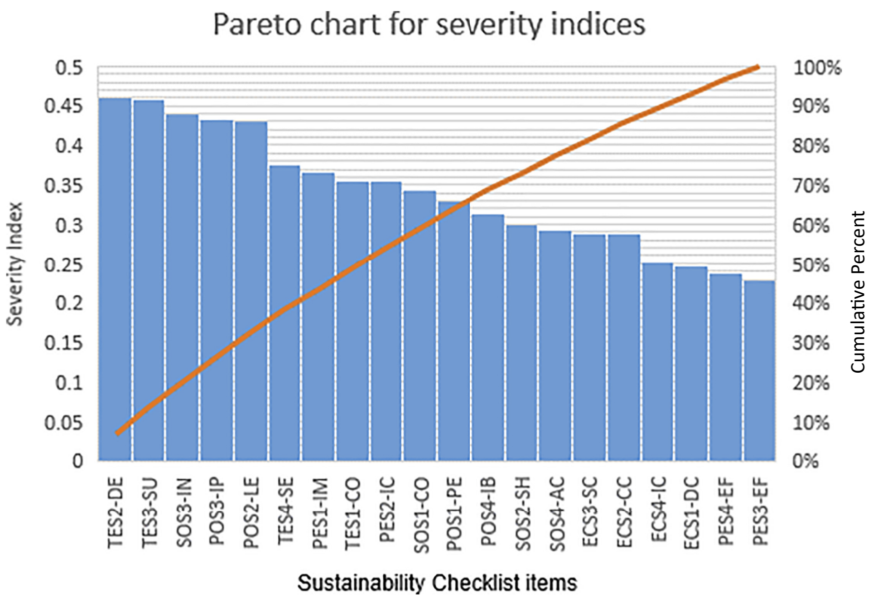

Figure 7.
Pareto chart for severity indices of the sustainability checklist items (input variables)

first ten items that cover almost 68 cumulative percentages for further improvement in the sustainability of the system, i.e. TES2-DE to SOS1-CO. Hence, administrators can handle a manageable number of sustainable factors for improving the overall sustainability of the system; conversely, original "80/20" rule recommended almost similar to the total number of sustainability factors considered. So, administrators can start improving sustainability from the factors beginning from TES2-DE to SOS1-CO, as mentioned in Figure 7. Also, system administrators or the management responsible for improving the sustainability of the system can solve each problem related to sustainable factors identified using the proposed method as availability of their time, funds, resources, etc.

According to Figure 5, conventional questionnaire-based sustainability methods considered the lowest mean scored variables to improve the overall sustainability of the system. In our study ECS2-CC, TES4-SE, TES3-SU, POS1-PE, ECS1-DC are the highest prioritized sustainability required factors. But, this recommendation is only based on the unfair lower user evaluation for a particular sustainability factor. In our proposed method, not only these lower evaluated factors but also each factor's impact on overall sustainability such as the severity index was considered. This is a fair measure of sustainability because it considers the individual impact of each sustainability factor performance within the machine learning prediction model. Therefore, the first iteration of our proposed sustainable evaluation algorithm recommended the highest impacted sustainability factors to be considered are TES2-DE, TES3-SU, SOS3-IN, POS3-IP, POS2-LE.

7. Conclusion

The main objective of this study is to evaluate the sustainability of the proposed ML system for higher education, which was developed based on the proposed ML framework for higher education. For that, the machine learning ensemble classifier-based novel sustainability evaluation approach was proposed. Furthermore, in this study, the majority voting classifier was used to identify the hidden relationship between the input variables and the output variable. The ensemble classifier consisted of three base machine learning models, i.e. DT, SVM and ANN. In total, 150 students and 150 teachers in the university community participated in this survey. The survey was used in the questionnaire with 20 questions. These 20 questions denoted 20 sustainability factors and categorized them under five main sustainability dimensions, i.e. economic sustainability, social sustainability, political sustainability, technological sustainability and pedagogical sustainability. According to the proposed sustainability evaluation algorithm and ensemble machine learning approach, practitioners are able to improve the overall sustainability of the system iteratively. In the first iteration, the results revealed that technological sustainability is questionable due to insufficient device availability and fewer supports from the technical staff. Also, social sustainability needs to be enhanced by improving community influence for the system. Political sustainability is also required to have more supportive institutional ML policy and strong administrative leadership. By addressing these sustainability issues, practitioners able to evaluate the system again and identifying further sustainability issues of the system.

8. Future improvements

In this study, a sample from one institute was selected. By taking the sample from diverse higher education institutes in various parts of the country will receive more responses to having a better sustainable ML framework.

References

Abdella, G.M., Kucukvar, M., Onat, N.C., Al-Yafay, H.M. and Bulak, M.E. (2020), "Sustainability assessment and modeling based on supervised machine learning techniques: the case for food consumption", *Journal of Cleaner Production*, Vol. 251, pp. 119661-119677.

Alharthi, A.D., Spichkova, M. and Hamilton, M. (2018), "Sustainability requirements for eLearning systems: a systematic literature review and analysis", *Requirements Engineering - Springer*, Vol. 24 No. 4, pp. 523-543.

ASCE (2018), "Policy Statement 818-The role of the civil engineer in sustainable development", available at: https://www.asce.org/issues-and-advocacy/public-policy/policy-statement-418--the-role-of-the-civil-engineer-in-sustainable-development/ (accessed 27 November 2020).

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2013), *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences*, Routledge, New Jersey.

Coskun-Setirek, A. and Tanrikulu, Z. (2019), "Technological sustainability of mobile learning", *The Online Journal of Science and Technology*, Vol. 7 No. 3, pp. 89-97.

Dietterich, T.G. (2000), *Ensemble Methods in Machine Learning*, Springer, Berlin, Heidelberg, pp. 1-15.

Dolawattha, D.D.M., Premadasa, H.K.S. and Jayaweera, P.M. (2019), "The impact model: teachers' mobile learning adoption in higher education", *International Journal of Education and Development using Information and Communication Technology*, Vol. 15 No. 4.

Harvey, H.B. (2018), "The pareto principle", *Journal of the American College of Radiology*, Vol. 15 No. 6, p. 931.

Hays, J. and Reinders, H. (2020), "Sustainable learning and education: a curriculum for the future", *International Review of Education*, Vol. 66 No. 1, pp. 29-52.

Jung, Y. (2018), "Multiple predicting K-fold cross-validation for model selection", *Journal of Nonparametric Statistics*, Vol. 30 No. 1, pp. 197-215.

Kang, J., Park, Y.J., Lee, J., Wang, S.H. and Eom, D.S. (2017), "Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems", *IEEE Transactions on Industrial Electronics*, Vol. 65 No. 5, pp. 4279-4289.

Keller, J.M., Liu, D. and Fogel, D.B. (2016), *Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation, s.l*, John Wiley & Sons, Hoboken, New Jersey.

Luan, W., Lu, L., Li, X. and Ma, C. (2018), "Integrating extended Fourier amplitude sensitivity test and set pair analysis for sustainable development evaluation from the view of uncertainty analysis", *Sustainability*, Vol. 10 No. 7, p. 2435.

Mabila, J., Biljon, J.V. and Herselman, M. (2017), "A sustainability framework for mobile technology integration in schools: the case of resource-constrained environments in South Africa", *The Journal of Community Informatics*, Vol. 13 No. 2.

Makridakis, S., Wheelwright, S. and Hyndman, R.J. (1998), *Forecasting Methods for Managers*, John Wiley & Sons, New York.

Moodle (2020), "Moodle app - MoodleDocs", available at: https://docs.moodle.org/310/en/Moodle_app (accessed 1 December 2020).

Ng, W. and Nicholas, H. (2013), "A framework for sustainable mobile learning in schools", *British Journal of Educational Technology*, Vol. 44 No. 5, pp. 695-715.

Nilashi, M., Rupani, P.F., Rupani, M.M., Kamyab, H., Shao, W., Ahmadi, H., Rashid, T.A. and Aljojo, N. (2019), "Measuring sustainability through ecological sustainability and human sustainability: a machine learning approach", *Journal of Cleaner Production*, p. 118162.

Nosratabadi, S., Mosavi, A., Keivani, R., Ardabili, S. and Aram, F. (2019), *State of the Art Survey of Deep Learning and Machine Learning Models for Smart Cities and Urban Sustainability*, Springer, Cham, pp. 228-238.

Ofei-Manu, P. and Didham, R.J. (2018), "Identifying the factors for sustainability learning performance", *Journal of Cleaner Production*, Vol. 198, pp. 1173-1184.

Pereira, F.C. and Borysov, S.S. (2019), "Machine learning fundamentals", *Mobility Patterns, Big Data and Transport Analytics*, s.l., Elsevier, pp. 2-29.

Pouezevara, S., Mekhael, S.W. and Darcy, N. (2014), "Planning and evaluating ICT in education programs using the four dimensions of sustainability: a program evaluation from Egypt", *International Journal of Education and Development Using Information and Communication Technology (IJEDICT)*, Vol. 10 No. 2, pp. 120-141.

Principe, J.C., Xu, D., Fisher, J. and Haykin, S. (2000), "Information theoretic learning", *Unsupervised Adaptive Filtering*, Vol. 1, pp. 265-319.

Principe, J.C., Euliano, N.R. and Lefebvre, W.C. (2001), *Neural and Adaptive Systems: Fundamentals through Simulations*, Wiley, New York.

Raschka, S. and Mirjalili, V. (2017), *Python Machine Learning Second Edition*, Packt Publishing Ltd, Birmingham.

Saltelli, A. (2002), "Making best use of model evaluations to compute sensitivity indices", *Computer Physics Communications*, Vol. 145, pp. 280-297.

Vaughn, P. (2018), *Fundamentals of Machine Learning*, Larsen and Keller Education, New York.

Venugopal, V. and Saleeshya, P.G. (2019), "Manufacturing system sustainability through lean and agile initiatives", *International Journal of Sustainable Engineering*, Vol. 12 No. 3, pp. 159-173.

Ziemba, E. (2017), "The contribution of ICT adoption to the sustainable information society", *Journal of Computer Information Systems*, Vol. 59 No. 2, pp. 116-126.

Corresponding author
H.K. Salinda Premadasa can be contacted at: salinda@ccs.sab.ac.lk