International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

A Novel Hybrid Machine Learning Model for Analyzing E-Learning Users' Satisfaction

Sulis Sandiwarno, Zhendong Niu & Ally S. Nyamawe

To cite this article: Sulis Sandiwarno, Zhendong Niu & Ally S. Nyamawe (2023): A Novel Hybrid Machine Learning Model for Analyzing E-Learning Users' Satisfaction, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2209986

To link to this article: https://doi.org/10.1080/10447318.2023.2209986

Published online: 28 May 2023.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2209986

A Novel Hybrid Machine Learning Model for Analyzing E-Learning Users' Satisfaction

Sulis Sandiwarno[a,b], Zhendong Niu[a], and Ally S. Nyamawe[c]

[a]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China; [b]School of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia; [c]Department of Computer Science and Engineering, The University of Dodoma, Dodoma, Tanzania

ABSTRACT
Analyzing lecturers' and students' satisfaction with using e-learning is important to improve the teaching-learning processes. The existing approaches have been widely employing machine learning algorithms, usage-based, and System Usability Scale (SUS) metrics based on users' opinions, activities, and usability testing, respectively. However, the usage-based and SUS metrics fail to cover users' opinions about e-learning systems and they involve manual features engineering. Whereas, the machine learning classifiers do not analyze satisfaction based on activities and usability. Toward this end, we propose a machine learning model that employs CNN and BiLSTM algorithms to concatenate the features extracted from users' activities, usability testing, and users' opinions. The proposed model is coined as E-learning Users' Satisfaction Detection (El-USD). Experimental results suggest that there is a significant correlation between satisfaction analysis by achieving an average $r = 0.778$. The evaluation results further suggest that our proposed approach can analyze users' satisfaction accurately.

KEYWORDS
E-learning; users' satisfaction; usage-based metrics; SUS; machine learning algorithms

1. Introduction

Recent, advancement in digital technology has significantly transformed education processes through improving learning and teaching (Almaiah, Alhumaid, et al., 2022). The use of the internet and computer technology has significantly changed the teaching and learning processes between lecturers and students from the conventional to the modern learning through electronic learning (e-learning). E-learning is a tool that delivers a collaborative means for achieving the knowledge and interactions between users, such as online forum, supporting of instructional and technical and quizzes without restriction of the time and space (Al-Fraihat et al., 2020; Almaiah, 2020; Almaiah, Hajjej, Lutfi, Al-Khasawneh, Shehab, et al., 2022; Duin & Tham, 2020; Girish et al., 2022; Hong et al., 2017; Sabeh et al., 2021; Tarus et al., 2018; Yilmaz, 2017; Zhu et al., 2023). E-learning systems are mostly used by institutions or universities and other educational organizations to provide new and innovative ways to deliver education to students (Shahzad et al., 2021; Wan & Niu, 2020).

To facilitate the learning processes, several Learning Management System (LMS) platforms have been proposed. LMS are education platforms based on web, cloud, or installed software that assist the learning processes and help effective instructions delivery (Pinho et al., 2021). Modi and Chaubey argued that other usual names in higher education, which are adopted interchangeably in place of LMS platform are Personal Learning Environment (PLE), Course Management System (CMS), and E-learning Courseware (EC) (Modi & Chaubey, 2021). Moodle is a common LMS platform mostly used to assist users in sharing materials and knowledge (Alkhateeb & Abdalla, 2021; Giannakos et al., 2022; Hasan et al., 2019; Perišić et al., 2018; Zabolotniaia et al., 2020).

Evaluation of e-learning systems has recently received considerable attention from the previous research to evaluate and quantify the satisfaction level of both lecturers and students. Satisfaction is a condition where the users accept the system and agree that they are comfortable to use it (Kornpitack & Sawmong, 2022; Landrum et al., 2021; Puška et al., 2021; Wei & Chou, 2020). In analyzing e-learning users' satisfaction based on users' activities, most of the previous approaches have attempted to employ several techniques, usage-based metrics, such as completion rate, task duration, and cursor distance or mouse click and most of the usability measurement tools for usability testing in analyzing e-learning users' satisfaction in the previous research is System Usability Scale (SUS) (Alghabban & Hendley, 2022; Harrati et al., 2016; Pal & Vanijja, 2020; Revythi & Tselios, 2019; Vlachogianni & Tselios, 2022). System Usability Scale (SUS) is a survey-based metric for assessing the perceived usability of a system, product, and service (Brooke, 1996). The SUS metric is suitable for analyzing users' satisfaction in e-learning systems since it can be used for a wide range of users' characteristics and it allows processing usability results quickly and easily (Ani, 2020; Ani et al., 2019). Furthermore, the aforementioned approaches

CONTACT Zhendong Niu zniu@bit.edu.cn School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China

have used Technology Acceptance Model (TAM) and Structure Equation Model (SEM) based on questionnaires to analyze users' satisfaction with the teaching and learning processes (Pal & Vanijja, 2020; Puška et al., 2021).

On the other hand, most of the previous approaches have widely used sentiment analysis in analyzing users' satisfaction (Adinolfi et al., 2016; Barrón Estrada et al., 2020; Elia et al., 2019; Hew et al., 2020; Nguyen et al., 2019; Onan, 2020; Oramas Bustillos et al., 2019; Qi & Liu, 2021; Vargas-Calderón et al., 2020). Sentiment analysis is defined as the contextual mining of text to identify and analyze the subjective opinions of users (Bing Liu, 2010).

However, the usage-based and SUS metrics approaches fail to cover users' reactions or ignore their opinions about e-learning systems and they involve manual features engineering that is often tedious, and time consuming. Besides, most of the existing approaches analyze lecturer's and students' satisfaction separately despite the fact that they are both important actors in learning processes, and analyzing their satisfaction at time could provide more insightful findings. Almaiah et al. argued that one of the successful e-learning systems' foundation is the aspect of interactive between lecturers and students of the LMS platform (Almaiah et al., 2020; Almaiah, Al-Lozi, et al., 2021). Furthermore, the existing approaches that leverage sentiment analysis by using machine learning algorithms do not analyze users' satisfaction based on activities and usability testing that have been shown to be effective in inferring users' satisfaction. To address such problems, we propose a novel machine learning model called E-learning Users' Satisfaction Detection (El-USD) that combines CNN and BiLSTM algorithms for text opinion classification, the usability measurement tool for usability testing in users' satisfaction, and usage-based metrics for analyzing e-learning users' satisfaction based on activities. The main contributions of this article are summarized as follows:

- First, we propose a novel machine learning-based approach to automatically analyze e-learning users' satisfaction based on users' activities, usability testing, and opinions.
- Second, under the proposed approach, we employ Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) algorithms to respectively extract the features from users' activities, satisfaction of usability testing, and to capture and learn the contextual information of users' opinions. Then, we combine the final representation by concatenating the outputs of CNN and BiLSTM algorithms for further analysis of users' satisfaction. To the best of our knowledge, this study is the first attempt to analyze e-learning users' satisfaction based on combining users' activities, usability testing, and users' opinions.
- Third, we curated a new e-learning dataset coined as Moodle e-learning dataset (MOELD) to support analyzing users' satisfaction. The dataset can be publicly accessed at https://fasilkom.mercubuana.ac.id/moeld/.
- Fourth, we compared and examined the correlation between the results of users' activities, users' usability measurement, and opinions using machine learning classifiers on real dataset collected from a well-known e-learning system. The evaluation results indicate that there is a significant correlation between the two approaches.

The remainder of this article is structured as follows: In the related work section, we review the related works about e-learning users' satisfaction. In the proposed methodology section, we describe the proposed methodology of analyzing e-learning users' satisfaction. In the experiments and evaluation section, we present the experimental settings and discusses the experimental results. Finally, we conclude the article and highlight the future work.

2. Related work

In this section, we review the existing approaches relating to analyzing users' satisfaction based on usage-based and SUS metrics, review studies of e-learning systems, and further present some literature related to users' satisfaction evaluation based on machine learning classifiers.

2.1. Analyzing users' satisfaction based on usage-based and SUS metrics

Assessing users' satisfaction in using e-learning systems is important to improve the teaching-learning process. Satisfaction is related to users' reaction on how easy and comfortable when they are using an e-learning system. Wang and Zhou defined satisfaction as a comfort level of users about the different features of the e-learning system service that they received (Wang & Zhou, 2022). Almarashdeh and Periaiya and Nandukrishna defined satisfaction as the extent to which users trust the information technology available to fulfill their informational requirements (Almarashdeh, 2016; Periaiya & Nandukrishna, 2023). Users' satisfaction is an important factor to make a course successful on e-learning systems (Ifinedo et al., 2018; Kara et al., 2021; Kurucay & Inan, 2017; Lin et al., 2019; Yawson & Yamoah, 2020; Yuen et al., 2019). Simões employed heuristic evaluation to evaluate the interface design of system and SUS metric to analyze users' satisfaction of the e-learning system (Moodle) (Simões & de Moraes, 2012). Cohen and Baruth evaluated users' satisfaction based on an anonymous questionnaire and ANOVA analysis was performed to examine the significance of the difference among groups in online learning by their personality (Cohen & Baruth, 2017). Giray introduced a model to seek the understanding experience of learning Computer and Software Engineering undergraduate students. The data was evaluated using quantitative (questionnaire) and qualitative techniques (Giray, 2021).

On the other hand, Harrati et al. suggested that questionnaire metric as standalone assessment are not adequate. Therefore, they proposed several common usage-based metrics (completion rate, task duration, and cursor distance/mouse click) to be used in tandem with the SUS metric. Furthermore, the authors evaluated the correlation between the usage-based and the SUS metrics in analyzing users' satisfaction by using Pearson Correlation Coefficient (PCC). The results indicate that there is a strong correlation between usage-based metrics and SUS metric (Harrati et al., 2016).

Users' satisfaction analysis based on questionnaires and usage-based metrics may not to be enough, because they fail to cover the users' performance and ignore to analyze users' navigation on e-learning system. Consequently, He et al. in their work adopted lostness metric to evaluate satisfaction based on how users navigate on e-learning systems. The authors evaluated the performance of their approach by investigating the correlation between lostness, the completion rate, and task duration metrics (He et al., 2018).

2.2. Challenges of E-learning systems

In this section, we review TAM and SEM techniques to analyze the model of teaching and learning processes. For example, Pal and Vanijja analyzed the perceived usability in online learning of Microsoft Teams in India by adopting System Usability Scale (SUS) questionnaire and Technology Acceptance Model (TAM) (Pal & Vanijja, 2020). Almaiah et al. employed the model of TAM to understand the main determinants that affect the intention of students to employ the digital and mobile games. The results indicated that the intention of students to employ mobile games was positively affected by perceived ease of use and perceived usefulness (Almaiah, 2020). Puška et al. proposed a model to evaluate students' satisfaction in using e-learning system. In evaluating the students' satisfaction, the authors focused on computer and students metacognitive' abilities and setting the objective of learning. To conduct the study, they used random samples and questionnaires. The collected data from students were evaluated and confirmed by adopting analysis of confirmative factors. Moreover, the Structural Equation Model (SEM) was employed to test the proposed model (Puška et al., 2021). Almaiah et al. presented a model of an assessment of an initiative that seeks to transcend the digital information technologies in higher education by using an integrative approach that quantifies both digital information flow and the quality of tutors (Almaiah, Alhumaid, et al., 2022). Almaiah et al. presented the core Madrasati platforms' determinants in supporting learning and teaching process in Saudi Arabia. The authors used a common method (i.e., SEM) to test the hypotheses of the proposed model and PLS-SEM software to analyze the data. The results indicated that the Madrasati platform increased use in Saudi Arabia was primarily attributed to the qualities of system service and contents as well as technological infrastructure (Almaiah, Hajjej, Lutfi, Al-Khasawneh, Shehab, et al., 2022).

On the other hand, the aforementioned approaches have been widely used TAM and SEM via Mobile learning (M-learning) applications in supporting educational processes (Almaiah, Al-Khasawneh, et al., 2021; Almaiah, Al-Otaibi, et al., 2022; Almaiah, Ayouni, et al., 2022; Althunibat et al., 2021). The results represented that the relationship between M-learning and IT infrastructure is positively accepted. Almaiah et al. argued that the M-learning application offers the mobility and flexibility benefits as a novel idea in contemporary education and an e-learning extension. Moreover, the authors hope that the M-learning will be suitable value to support the learning processes (Almaiah, Hajjej, Lutfi, Al-Khasawneh, Alkhdour, et al., 2022). Moreover, Almaiah et al. introduced a new model to identify the crucial factors that influence students' adopt the platform of mobile learning during the COVID-19 pandemic. The authors employed the modeling technique of ANN and SEM to analyze the fundamental relationship among paradigms in the model of research. The experimental result indicated that the mobile learning platform is important to conduct the processes of learning and teaching (Almaiah, Al-Lozi, et al., 2021).

Although, the aforementioned approaches achieved good results, but the usage-based and SUS metrics using manual feature engineering and is tedious and time consuming. Furthermore, they ignore the assessment of users' reactions or opinions which are potential in analyzing e-learning users' satisfaction levels. Qika et al. and Onan in their work argued that questionnaire scores are not sufficient measures to express the true acceptance of an e-learning system by users (Lin et al., 2019; Onan, 2020). Therefore, a possible solution is to include the assessment of users' reactions or opinions, usability testing, and users' activities through the use of sentiment analysis and machine learning techniques.

2.3. Machine learning-based classifiers

Recently, machine learning-based classifiers have been successfully adopted in sentiment analysis and on other fields. Khan et al. defined sentiment analysis as a process to determine whether the polarity of a textual corpus (i.e., sentence, paragraph, document, etc.) is positive, negative, or neutral (Khan et al., 2017). Kontopoulos et al. argued that sentiment analysis is a growing field in the Natural Language Processing (NLP) research community that focuses on text mining and computational linguistic to identify the opinions of users about a service or system (Kontopoulos et al., 2013). Additionally, sentiment analysis or opinion mining is used for identifying the sentiment orientation and assess the users' emotional strength (Feng et al., 2011; Onan et al., 2016). Generally, sentiment analysis focuses on text extraction and text classification aspects.

Text classification nowadays is the most popular technique in machine learning, which involves assigning an unknown textual document to appropriate classes. For example, Kechaou et al. proposed two machine learning algorithms to evaluate the e-learning system based on lecturers' opinions. In their work, they adopted Hidden Markov Models (HMM) and SVM algorithm to classify the opinions. Furthermore, they employed several methods of feature selection, such as Information Gain (IG), Mutual Information (MI), and CHI statistics (CHI). The results indicate that SVM with IG outperforms other classifiers

(Kechaou et al., 2011). Adinolfi et al. introduced the framework of sentiment analysis for the purpose of higher education evaluation and showing how it is adopted to measure the satisfaction of students against lecturers' performance on online platforms, such as Massive Open Online Course (MOOC) (Adinolfi et al., 2016). Maitra et al. proposed a machine learning model to evaluate the students' opinions on the learning process for helping the faculty based on Naïve Bayes (NB) classifier. The results indicate that the NB classifier is effective in evaluating students' opinions (Maitra et al., 2018). Skrbinjek and Dermol focused on predicting students' satisfaction and how to relate satisfied students to their performance in the classroom. The results revealed that users were less satisfied with a course when the requirements for involvement and the work-load in the classroom were both high (Skrbinjek & Dermol, 2019). Elia et al. proposed a novel model to measure lecturers' satisfaction against the course by adopting an analysis of Big Data. They suggested the Big Data as the right paradigm for real time processing large datasets regarding the users' satisfaction in the environment of collaborative learning. Moreover, the authors presented a software artifact that included the Big Data Learning Analytics (BD-LA) regarding real-time perceptions of the evaluation strategy of online courses (Elia et al., 2019). Hew et al. employed supervised machine learning classifiers to analyze the features and students' perceptions of MOOC. The results indicated that the machine learning classifiers can significantly predicted the users' satisfaction (Hew et al., 2020). Liu and Yang proposed a machine learning model and high-frequency analysis of the lexical method to determine the factors which affect the MOOCs' development by analyzing the negative sentiments of students' feedback. To analyze the students' feedback, the authors propose to use several machine learning classifiers, such as KNN, SVM, LR, and NB (Liu & Yang, 2020). Giang et al. introduced a system that enables the sentiment classification of students' feedback to analyze users' satisfaction in online learning. The authors collected students' feedback data which contains 5,000 sentences and then annotated with three labels that is positive, negative, and neutral. The experimental results suggest that Maximum Entropy (ME) classifier outperformed NB and SVM classifiers (Giang et al., 2020). Ho et al. investigated the important predictors in determining the undergraduate students' satisfaction from multiple departments in adopting Emergency Remote Learning (ERL) where Moodle platform and Microsoft Team are the key tools of learning. In supporting the study, the authors proposed to use several machine learning classifiers, such as KNN, SVM, MLPR, LGBM, and RF. By comparing the multiple regression and machine learning classifiers before and after the use of random forest elimination of recursive feature, all multiple regression and machine learning classifiers attained the improved accuracy (Ho et al., 2021).

On the other hand, several previous approaches have attempted to use deep learning-based techniques to analyze the users' opinions in educational domain. For instance, Nguyen et al. adopted conventional machine learning and deep learning classifiers on the corpus of Vietnamese students' opinions. The results show that deep learning techniques (i.e., LSTM and BiLSTM) outperformed two machine learning classifiers (NB and ME) (Nguyen et al., 2019). Oramas et al. in their study presented deep learning techniques for opinions' analysis in an intelligent learning environment. In the presented schemes, several conventional machine learning classifiers, such as Bernouli NB, Multinomial NB, SVM, linear SVM, KNN, and Bayesian Network have been adopted. Whereas, in the deep learning techniques, they proposed to use CNN and LSTM algorithms. After training the dataset, the results indicate that deep learning techniques outperformed conventional machine learning classifiers (Oramas Bustillos et al., 2019). Onan in their study proposed to use several conventional machine learning (i.e., NB, SVM, LR, KNN, and RF) and several deep learning algorithms, such as CNN and LSTM to evaluate instructors' evaluations. Additionally, the author employs some three conventional text representation schemes, such as Term Frequency (TF), Term Presences (TP), and Term Frequency Inverse Document Frequency (TF-IDF), and word embedding techniques, such as Word2Vec, GloVe, and FastText (Onan, 2020). On the other hand, the aforementioned approaches have proposed to use Bidirectional Encoder Representation Transformers (BERT) (Jafarian et al., 2021; Li et al., 2021; Pota et al., 2020; Wang et al., 2020) algorithm and obtained better performance in the field of sentiment classification.

In summary, the aforementioned approaches only analyze the satisfaction of lecturers and students separately despite the fact they both play a critical role in e-learning systems, and they also ignore to evaluate the level of users' satisfaction by tasks. In addition, to the best of our knowledge, none of the deep learning approaches have attempted to analyze e-learning users' satisfaction based on users' activities, usability testing, and users' opinions.

3. Proposed methodology

The main objective of this study is to analyze users' satisfaction in e-learning systems. As depicted in Figure 1, we describe the framework of the proposed methodology for analyzing users' satisfaction. The framework has three steps as follows: First, we collect users' activities, usability testing, and opinions from an e-learning system. Second, we train to classify the users' activities, usability testing in the tasks, and the users' opinions by using CNN and BiSLTM classifiers. Third, we concatenated the CNN and BiLSTM results for representing the satisfaction levels of all sub-tasks synchronously. The following subsections present in detail each of the key steps of our approach.

3.1. E-learning users' activities and usability measurement

A tree-based graphical representation is proposed for building a task model that describes the tasks, actions, and goals to be achieved by a participant (i.e., students or lecturers).

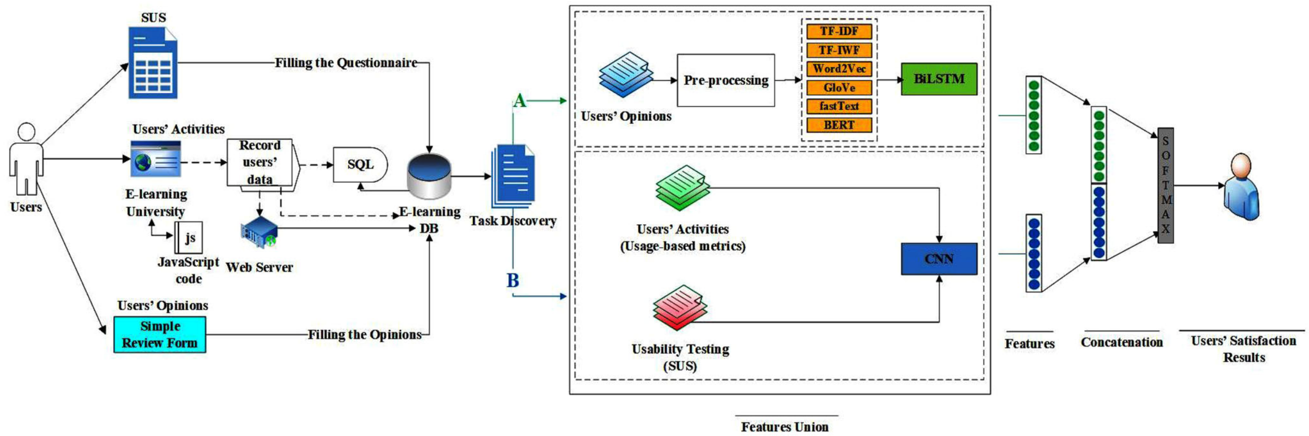

Figure 1. Framework of the proposed methodology.

The resulting task model tree represents all lecturers' and students' interactions that they can perform on a given web interface. Using a visual notation of the tree-based model, the task model is a reserved hierarchy of tasks to be completed to accomplish a specific goal of the task. A task that is reflected only by name consists of users' actions to be accomplished. This can be a basic task consisting mainly of simple users' actions, such as clicking a submit button, scrolling of a web page, and typing text into a text field. For the context of this study, Figure 2 describes the five task models of users for using e-learning system (i.e., Moodle). To acquire data on the courses in the e-learning system, we collected from the data logs of the e-learning systems and put-on JavaScript-code into the e-learning systems for collecting data on activities performed by the lecturers and students in the course. The events recorded by JavaScript to assess users' satisfaction based on system usage, we define the series of tasks (task descriptor), which are generally conducted by lecturers and students in e-learning system.

The task descriptors of lecturers and students are similar but some tasks are different. The task model contains five consecutive high-level tasks that reflect the users' activities, such as login, browsing the course, creating the discussion forums, replying discussion forums, and uploading quizzes by lecturers for their students.

During the first task, lecturers and students are presented with the front page containing a link of login at the top header that would take the users to the login page with a form asking the users to fill their credentials. Second, users would start the second task which is browsing the course and permitted to the editing mode. Third, users are required to open the forum page, filling the form, and submit it. Fourth, users are required to respond in the discussion forums. Finally, lecturers are asked to create questions by filling in the form of e-learning system. Afterwards, they click on the submit form button to complete the quiz session. Note that, all the lecturers' and students' actions are recorded automatically by logging into the database. Moreover, to further asses the usability of the e-learning system, the users had to fill out a 10-questions SUS questionnaire. The questionnaire was based on the Likert scale with values ranging from 1 up to 5. Consequently, the responses were used to deduce the SUS to quantify the usability of the e-learning systems. In supporting the evaluation of users' satisfaction based on activities in e-learning system, we employ several algorithms, such as SUS, completion rate, task duration, and lostness.

1. Task Duration is the time elapsed from the start of a task to its completion, usually expressed in minutes and seconds (Albert et al., 2010; Tullis & Albert, 2013).
2. Completion rate is a metric used to measure the percentages of task completion assigned to users (Harrati et al., 2016; Tullis & Albert, 2013). The completion rate typically assesses the percentage of task completion ranging from 0% (failure) to 100% (success). The success of each activity depends on the user's ability to complete the assigned task. If users are deemed incapable of completing a job, then the result will not be good. The completion rate metric is defined as:

$$C = \frac{\sum task_{complete}}{\sum task_{undertaken}} \times 100\% \tag{1}$$

3. Lostness is an effective metric for calculating navigation on web pages in which the participants took to complete the task step by step (Smith, 1996; Tullis & Albert, 2013). To calculate lostness, three values are used as proposed by (Smith, 1996), n represents the number of different web pages visited while performing the task, s is the total number of pages visited to perform each task, r denotes the minimum (optimum) number of pages that must be visited to accomplish the task. Suppose the user started down some incorrect paths before finally getting to the right place, visiting a total of different pages (n). Therefore, lostness L can be formally defined as:

$$L = \sqrt{\left(\frac{n}{s} - 1\right)^2 + \left(\frac{r}{n} - 1\right)^2} \tag{2}$$

4. SUS is one of the popular metrics used to measure the usability of systems and users' satisfaction (AlGhannam et al., 2018; Bangor et al., 2008; Brooke, 1996). SUS contains 10-questions with a mix of positive and negative questions. For each SUS question, users rate the

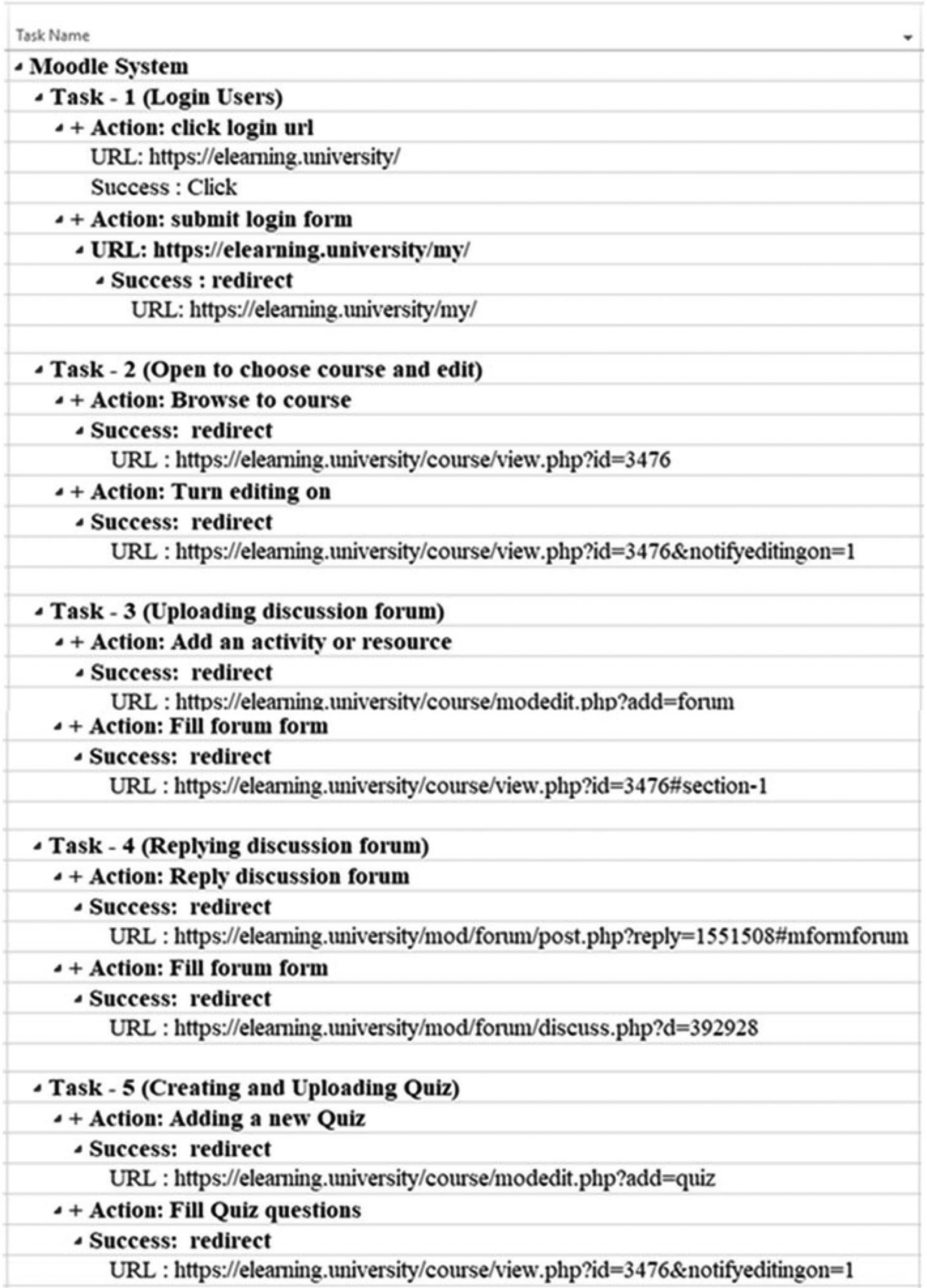

Figure 2. Five task descriptors for users in e-learning systems.

magnitude of their agreement employing 5-point Likert scale that ranges from 1 (disagree) to 5 (strongly agree). To calculate the SUS score, for each odd question that is worded positively the scale minus 1, and for each even question the score is 5 minus the position of a scale. SUS sums all scores for 10-questions multiplied by 2.5. The score of SUS ranges between 0 to 100 in 2.5-point increments where the top values indicate higher users' satisfaction. SUS can be formally defined as:

$$SUS = 2.5\left[\sum_{n=1}^{5}(U_{2n-1}-1)+(5-\ U_{2n})\right] \tag{3}$$

3.2. Task discovery in analyzing users' satisfaction

In recent years, the widespread application of machine learning classifiers has been successfully employed in the classification of users' opinions (Lin et al., 2019). Generally, to classify the data of users in the e-learning systems by using machine learning classifiers we follow these three key steps: First, we pre-process the data of users' opinions. Second, we create a feature modeling to extract the features. Finally, we use the extracted features to train the machine learning classifiers to analyze users' satisfaction. The following subsections present in detail each of these key steps.

3.2.1. Opinions pre-processing

Opinions pre-processing or text pre-processing is the key step that requires the cleaning and preparing the data for the classification, which consequently improves the performance of classification. The standard text pre-processing steps are applied to prepare the opinions data for classification. The goal is to transform the texts in opinions which are generally the free-form texts into a form that is suitable for textual analysis. We leveraged the Python Natural Language Toolkit (NLTK)[1] for such purpose. The applied pre-processing techniques include tokenization, stopwords removal, removing HTML tags, and removing bad characters (Madani et al., 2019; Umer et al., 2018). First, tokenization involved breaking up opinions' examples into a list of individual words (i.e., tokens). In this step, some characters, numbers, and punctuation are excluded as they do not contain any useful information. Second, stopwords removal is employed, the general and regularly applied words, such as "the," "a," "an," "it," "with," "and," "that," "which," etc. are removed as they do not carry any useful information and they just introduce noise. Third, removing HTML tags is used, the commonly applied tags, such as "http:/," "https:/," and "www" are eliminated as they are not considered important. Finally, removing bad characters is leveraged to further clean the opinions.

3.2.2. Feature modeling techniques

Computer classifiers can only work with the numerical data, where to process the data especially textual information in the computer, those data should be represented in format of numerical. In this study, we are adopting several methods that were motivated by previous work to represent the process of the text, such as (TF-IDF) (Salton, 1991), (TF-IWF) (Wang et al., 2008), Word2Vec (Mikolov, 2015), (GloVe) (Pennington et al., 2014), fastText (Bojanowski et al., 2017), and BERT (Patwa et al., 2020).

3.2.3. Machine learning classifiers

Machine learning classifiers have been widely used in a wide-range of text classification tasks, that involve statistical and probabilistic techniques. Friedman et al. argued that Probabilistic techniques are among the most popular classification models that they are based on the Bayes theorem for estimating the conditional probability of class label y (Friedman et al., 1997). For an input vector $y = \; y_i, \ldots, y_n \in \mathbb{R}^{document}$, the machine learning classifier aims at classifying y into a class in a discrete possible class $c \; = \; \{c_1, \ldots, c_n\}$. A model of machine learning classifier is defined as M and a classification function $C_M\colon \; \mathbb{R}^{document} \rightarrow \; \{c_1, \ldots, c_n\}$. To predict the class in which y belongs, the function $C_M(.)$ should be evaluated on y. The result of the classification is defined as $c_{i_o} \leftarrow C_m(y)$.

In this study, we propose a model of deep neural network to address the analyzing users' satisfaction in e-learning systems. The architecture of this model comprises a combination of two different deep neural networks, which are CNN and BiLSTM algorithms for extracting and capturing the long-term dependencies of the input users' activities, usability testing, and opinions, respectively. Yadav and Dinesh argued that the CNN classifier is one of the most deep learning architecture, which is a multi-layered feed forward model of neural network (Yadav & Vishwakarma, 2020). Generally, the CNN classifier consists of three layers, such as convolution, pooling, and fully connected layer as shown in Figure 3. The representation of a CNN classifier takes the data as an input and designates the relationship between the CNN model layers for determining the class of the data. In the layer of convolution, the features are extracted from the data with several filters help. The process of midway is used between the layers of convolution and pooling so that the features non-linear with the Rectified Linear Unit Activation function help. In the layer of pooling, the dimension of these feature maps are subtracted, which subtracts the convolution performance in the layers of subsequent and display the features in the data more effectively and efficiently.

Generally speaking, the LSTM classifier processes the representation of a sentence vector which takes as an input from the first to the last words. Whereas, the BiLSTM classifier is one of RNN classifiers to expand the LSTM classifier which has shortcoming of the text sequence features.

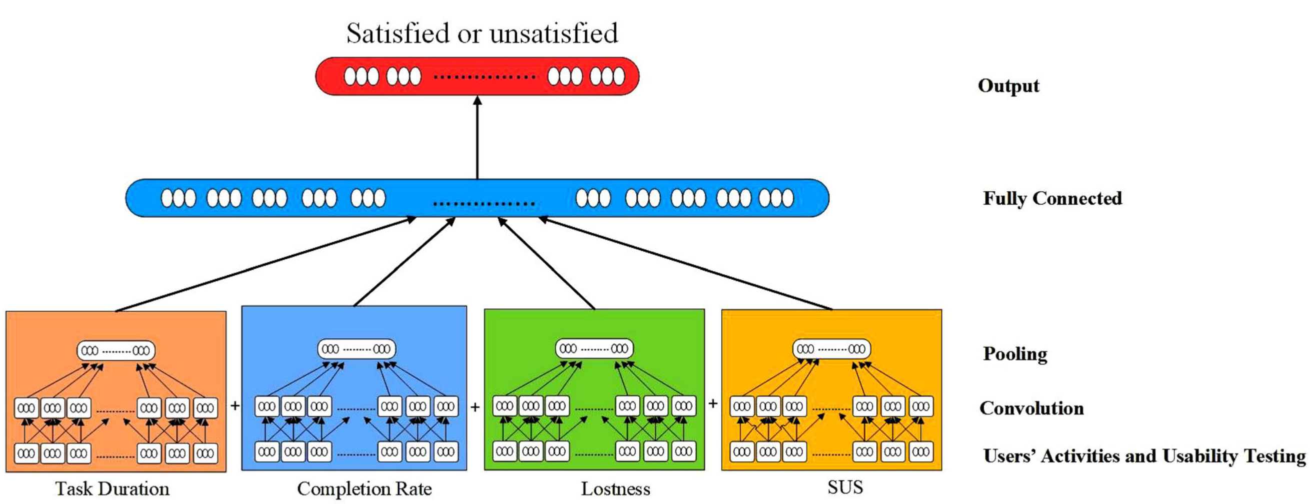

Figure 3. CNN architecture of analyzing e-learning users' satisfaction.

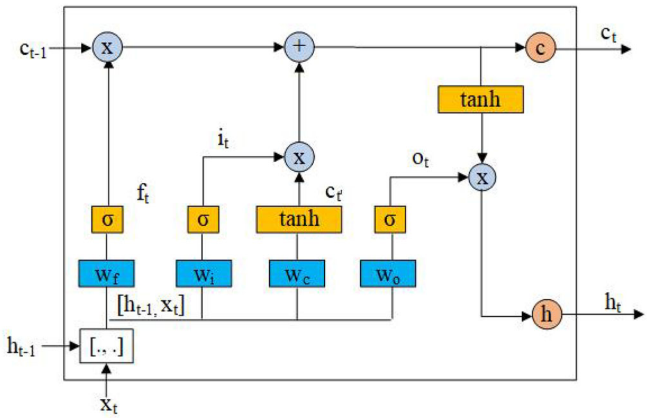

Figure 4. Architecture of BiLSTM.

Liu et al. and Niu et al. argued that the BiLSTM better than LSTM classifiers in solving the task of sequential modeling (Liu et al., 2017; Niu et al., 2017). This can help the BiLSTM classifier to learn and capture the context better. The node of BiLSTM classifier consists of three gates, such as input gate (notated as i_w), output gate (notated as o_w), and the forget gate (notated as f_w) as shown in Figure 4.

The C_{t-1} represents the previous moment cell state, h_{t-1} is defined as the BiLSTM neuronal unit final output value at the last moment, x_t denotes the input for the moment of current, σ is the logical function of Sigmoid, c_t is the cell state at the current moment, and the h_t is notated as the values of current latency. Which is, the process of BiLSTM calculation is as follows (Guo et al., 2018):

$$i_w = \sigma \left(w_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4}$$

$$f_w = \sigma \left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

$$o_w = \sigma \left(w_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{6}$$

$$c_t' = \tanh\left(w_c \cdot [h_{t-1}, x_t] + b_i\right) \tag{7}$$

$$c_t = f_w \times c_{t-1} + i_w \times c_t' \tag{8}$$

$$h_t = o_w \times \tanh(c_t) \tag{9}$$

where c is the Cell vectors. Which is, the dimension of these vectors is consistent with the hidden layer vector dimensions h. Whereas, w_i, w_f, w_o are represents the weights of input, forget, and output gates, b_i, b_f, and b_o are denotes bias of the input, forget, and output gates.

The BiLSTM classifier has been widely used frequently to identify the relationship of long-term among the words from the beginning to the end and reverse direction as shown in Equations (10) and (11).

$$\overrightarrow{h_t} = \overrightarrow{LSTM}\left(w_i, \overrightarrow{h_{t-1}}\right), \; t \in [1, \; N] \tag{10}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}\left(w_i, \overleftarrow{h_{t-1}}\right), \; t \in [1, \; N] \tag{11}$$

To increase the integrity of satisfaction level, the combination of different e-learning users' data will use to express the integrity of the users' satisfaction. In this study, the same users' data are represented by all feature extraction techniques. These different representations are the inputs of different deep learning classifiers and extracted the features. According to Figure 1, there are three steps in the model, the first step works with the BiLSTM classifier. The second step works with the CNN classifier. Then, we concatenate the results of CNN and BiLSTM classifiers.

In step A, the dataset D consisting of n users' e-learning is represented as $D = \{E_1, E_2, E_3, \ldots, E_n\}$, and n represent the equal to 74,050 total users' e-learning in the dataset D. Given a users' e-learning E_i, the users' e-learning with D words w_i is declared as $E_i = \{w_{i,1}, w_{i,2}, w_{i,3}, \ldots, w_{i,k}\}$. Where, in step A of the proposed method, we embed each word w_i of users' e-learning to a vector of pre-trained word. Which is, $w_l^{\omega} \in \mathbb{R}^d$ represents the d vector of dimensional embedding of the l^{th} word w_l^{w} and the embedding of word level is denoted as the $E_i^{\omega} = \{E_{i,1}^{\omega}, E_{i,2}^{\omega}, E_{i,3}^{\omega}, \ldots, E_{i,k}^{\omega}\}$. We fed the E_i^{ω} as the input to BiLSTM algorithm for feature extraction technique, specifically as shown in Equation (12).

$$G_i^{BiLSTM} = BiLSTM\ \left(E_i^{\omega}\right) \tag{12}$$

Whereas, in step B, we supposed that V is denoted as the vector of the classical model (i.e., users' activities), and the d represents the dimensionality of classical model. The size of vector V is denoted as S. The matrix of the classical model will be equal to $Y \in \mathbb{R}^{h \times L}$, where h is denoted as the height of Y embedding matrix and L represents the length. The $E_{i,j}$, which is a task that consists of a classical model sequence $E_{i,j} = \{c_1, c_2, c_3, \ldots, c_p\}$. Where p is the length of $E_{i,j}$, j represents j^{th} task in the E_i. The task of classical model $E_{i,j}$ is $C^{E_{i,j}} \in \mathbb{R}^{h \times p}$. For each extracted of E_i, G_i^{CNN} features with the CNN classifier help and Y matrix as defined in Equation (13).

$$G_i^{CNN} = CNN\ (E_i) \tag{13}$$

Finally, by concatenating $\oplus$ the features achieved from CNN and BiLSTM classifiers steps namely $[G_i^{CNN}, G_i^{BiLSTM}]$ as the full connection layer input, the users' satisfaction is perceived as shown in Equation (14).

$$G_i^{El-USD} = \sum_{i=1}^{n} G_i^{CNN} \oplus G_i^{BiLSTM} \tag{14}$$

The achieved features of G_i^{El-USD} are fed into softmax layer of proposed hybrid method. In this way, the features of high level are conveyed to the softmax layer. The main novelty in this study is that we combine the different techniques to achieve a better users' satisfaction score with distinctive deep learning-based classifiers.

The final polarity of predictive users' satisfaction of one aspect of the goal is simply the label with the greatest probability. We minimize the regularization of cross-entropy loss by using *loss* function to train our model

$$loss = -\sum_{i}\sum_{c \in C} I\ (b_i = c).\log\left(P(b_i = c)\right) + \lambda ||\vartheta||^2 \tag{15}$$

where, $I\ (.)$ denotes the function of indicator, λ represents the parameter of *loss* function regularization, and the ϑ is a set of the weight matrix of CNN and BiLSTM networks and linear layers. After that, we employ Dropout to reduce the overfitting by randomly ignoring the neuron in the structure of neural network (Srivastava et al., 2014). Moreover, we are using the algorithm of Adam to minimize the *loss* function of the weight matrix and term of bias in the proposed model.

4. Experiments and evaluation

In this section, we empirically investigate the performance of our approach and compare it against the baseline methods for analyzing users' satisfaction.

4.1. Experimental environment

In this study, we used an Intel core I7-9750H computer with 1TB internal memory, GTX 1660Ti, installed with Win10Home as the operating system and Python programming version 3.9 as the development environment. Moreover, we used a Python library named Scikit-learn[2] to implement the machine learning classifiers.

4.2. Evaluation measures for the machine learning classifiers

We evaluate the performance of each classifier based on 10-fold cross-validation and measure the performance in terms of macro-averaged precision (Pre), recall (Rec), and F1-score (F1) (Lauscher et al., 2017). The randomly distributed users' opinions (notated as *uo*) are split into 10 same-sized categories notated as $O_i(i = 1....10)$. For the *i*th cross-validation, we are considering all opinions except those in O_i as the corpus of the training dataset, and then treat the opinions in O_i as the testing dataset. For processing *i*th fold cross validation, we are extracting all opinions as op_train_i from the training dataset, this is the union of all opinions except O_i.

$$op\_train_i = \cup_{k\in[1,10]^k \neq i} O_i \tag{16}$$

On the other hand, we examined the correlation results between our proposed algorithm and common feature extraction techniques on the machine learning classifiers using Pearson Correlation Coefficient (PCC) algorithm. PCC is an important algorithm to assess the correlation between variables (Adler & Parmryd, 2010; Benesty et al., 2009). Generally, the PCC algorithm denotes the value between [−1, 1]. When the correlation coefficient values near equals 1 depicts a perfect positive relationship, −1 represents a perfect negative relationship, and 0 denotes the blank relationship absence between variables. Generally, the PCC method is defined as:

$$PCC = \frac{\sum_{i=1}^{n}(zx_i - \overline{zx})(zy_i - \overline{zy})}{\sqrt{\sum_{i=1}^{n}(zx_i - \overline{zx})^2 (zy_i - \overline{zy})^2}} \tag{17}$$

where, $\overline{zx}$ represents the mean of zx variables and $\overline{zy}$ is defined as the mean of zy variable. Moreover, the deep learning parameter settings in the experiment are shown in Table 1.

4.3. Evaluation measures for the usage-based and SUS metrics

For the evaluation, we compared the proposed method against the traditional approaches. Note that, the completion rate results indicate that users are satisfied in the range of 70–100% (Albert et al., 2010; Harrati et al., 2016; Tullis & Albert, 2013). The result of lostness metric should be <0.5 to consider satisfaction (Smith, 1996; Tullis & Albert, 2013). The SUS results indicate that users are satisfied if the score is not <70 (AlGhannam et al., 2018; Tullis & Albert, 2013). Furthermore, we use Cronbach's Alpha motivated by Borkowska and Jach to evaluate the reliability of SUS scores. The Cronbach's Alpha is a metric that used to measure the reliability of a scale set or items of test. Cronbach's Alpha is calculated by correlating the score for each item of scale with the total score for each observation. After that, comparing that to the variance for all individual scores of items (Borkowska & Jach, 2017). The internal consistency measure of the scale in Cronbach's Alpha should reach the value above 0.8 (0.805). The Cronbach's Alpha metric is computed as:

$$\alpha = \frac{b}{b-1}\left[1 - \frac{\sum_{i=1}^{b}\sigma_{x_i}^2}{\sigma_y^2}\right] \tag{18}$$

where b represents the scale items number, $\sigma_{x_i}^2$ denotes the associated of variance with item i, and σ_y^2 is the associated of variance with the observed total scores.

4.4. The partial least squares-structural equation modeling

In this study, the analysis of data was conducted adopting the partial least squares-structural equation modeling (PLS-SEM) through SmartPLS (V. 3.2.9). The collected data was analyzed by employing a two-step measurement approach, which includes the measurement and structural models (Al-Emran & Salloum, 2017). The PLS-SEM was selected in this study for numerous factors. Firstly, if the given research aims to work on a current theory, the preference should be given to PLS-SEM (Lutfi, 2022). Second of all, the PLS-SEM can help with effectively handling exploratory research that has complex models (Almaiah, Al-Otaibi, et al., 2022). Third, PLS-SEM carries out an analysis of the entire models as one unit rather than making subdivisions out of it (Almaiah et al., 2016). Fourth, PLS-SEM also provides analysis of concurrent for the structural and models of assessment, since which accurate assessments are generated (Almaiah, Alfaisal, et al., 2022).

Table 1. The parameters' setting of deep learning classifiers.

Parameters' setting	CNN	BiLSTM
Learning rate	0.001	0.001
Batch size	128	128
Dropout	0.5	0.5
Number of filters	–	100

Table 2. The distribution of lecturers' and students' dataset.

Users	Fields	Total	(%)
Lecturers (L)	Computer Science (CS)	95	19
	Management (Mn)	207	41.4
	Engineering (En)	60	12
	Accounting (Ac)	138	27.6
Students (S)	Computer Science (CS)	2745	19.18
	Management (Mn)	5965	41.68
	Engineering (En)	1702	11.89
	Accounting (Ac)	3898	27.24

4.5. Data collection

The dataset in this study was collected from Universitas Mercu Buana, Jakarta, Indonesia.[3] The dataset contains data from various domains including Computer Science (CS), Management (Mn), Engineering (En), and Accounting (Ac). Note that, the dataset is comprised of two typical sets (i.e., users' activities and users' opinions). The total number of users for each set is 14,810 of which 500 are lecturers and 14,310 are students. Table 2 shows the distribution of the dataset. For training the machine learning classifier, the users' opinions can be provided with a 3-point scale score, from which an overall quality scores have been computed for a course. To obtain a labeled corpus of users' opinions we have adopted the quality scores supplied by users. In the evaluation opinion with the quality scores of -1 have been labeled as negative, 0 have been labeled as neutral, and 1 have been labeled as positive opinions.

4.6. Results and discussion

Table 3 depicts the classification results based on TF-IDF, TF-IWF, Word2Vec, GloVe, fastText, and BERT algorithms in terms of average macro-precision (Pre), recall (Rec), and F1-score (F1). From Table 3 we observe that based on TF-IDF algorithm the machine learning classifiers can accurately classify users' opinions with an average at least 62.64% of F1. Whereas, based on Word2Vec the classifiers can accurately classify opinions with an average at least 68.09% of F1. In addition, El-USD using Word2Vec outperforms TF-IWF algorithm with an average improvement of 2.05, 2.03, and 2% in terms of precision, recall, and F1, respectively. Broadly speaking, TF-IDF has been shown to be effective and slightly improves the classification performance. Additionally, TF-IDF algorithm can obtain good performance on short textual dataset and it can mark the importance of word w_i (Hasan & Ng, 2014). However, the limitation of TF-IDF is that it ignores the importance of Word Frequency (WF) in the text. To address the TF-IDF limitation, Wang et al. proposed TF-IWF algorithm. TF-IWF algorithm influences the WF reciprocal to replace the Inverse Document Frequency (IDF) with Inverse Word Frequency (IWF) (Wang et al., 2008). Although, the TF-IWF algorithm is the conventional feature extraction technique, it can successfully employ document frequencies to compute the normalized frequency in view of a corpus. Therefore, to support in investigating users' satisfaction in e-learning system based on opinions, the TF-IWF algorithm outperforms TF-IDF feature extraction techniques.

MNB is another algorithm that has shown to be effective as well in multi-class classification. Furthermore, the difference score between MNB and DT classifiers using fastText algorithm has been reported as 4.92, 4.94, and 5.03% in terms of precision, recall, and F1, respectively. We further note that MNB outperforms SVM, DT, LR, and KNN algorithms in both TF-IDF, TF-IWF, Word2Vec, GloVe, fastText, and BERT. Based on the preceding analysis we conclude that MNB algorithm is more effective than other classifiers. According to Kim et al. argued that MNB classifier has several superior advantages compared with other classifiers in classifying the text classification even though this classifier is old (Kim et al., 2018). Additionally, Jiang et al. argued that MNB algorithm is an efficient classifier and can scale well in several fields, such as text classification (Jiang et al., 2019).

On the other hand, the CNN algorithm based on fastText improves precision, recall, and F1 by 2.44, 2.36, and 2.43%, respectively compared to CNN based on GloVe algorithms. Compared with the CNN based on TF-IDF and the CNN based on fastText superior in its performance and reported an improvement of 5.05, 5.24, and 5.17% in precision, recall, and F1, respectively. El-USD based on fastText algorithm compared to the CNN based on fastText algorithm improves the results by 8.93, 8.96, and 8.97% in precision, recall, and F1. Furthermore, El-USD based on fastText algorithm compared to the CNN + LSTM based on fastText algorithm improves the results by 2.94, 2.93, and 2.94% in precision, recall, and F1. Moreover, El-USD based on BERT algorithm compared to the El-USD based on fastText algorithm improves the results by 6, 5.95, and 5.97% in precision, recall, and F1. It is noted that our proposed El-USD algorithm based on BERT outperforms and improves other machine learning classifiers as shown in Figure 5.

Although, several approaches have attempted to use deep neural network based on BERT algorithm and argued that this algorithm has problems, such as limited to deliver and generalized on several tasks in sentiment information (Batbaatar et al., 2019). Furthermore, this algorithm recent models of pre-training are often heavily data-driven, which is not reasonable to solve the problems with the comparatively small dataset (Zhang et al., 2021). In this study, BERT algorithm outperformed in each classifier for analyzing e-learning users' satisfaction.

Among the compared configurations, the highest predictive performance has been attained by El-USD in conjunction with BERT algorithms, with an average classification F1 of 93.08%. Generally, the results suggest that there is no significant difference in the classifiers individual performance across the folds as shown in Figure 5.

4.7. Parameters' impact of machine learning classifiers

In this section, we would like to evaluate the impact of the parameters of the proposed model (El-USD) and compared it with the other machine learning-based classifiers. Generally, we consider the following parameters, such as batch size, dropout, learning rate, and number of filters as

Table 3. The performance of machine learning classifiers results using six feature extraction techniques.

		Positive (%)			Negative (%)			Neutral (%)		
Models	Metrics	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
TF-IDF	LR	58.31	58.62	58.39	55.26	55.31	55.37	53.33	53.28	53.33
	KNN	59.46	59.21	59.44	55.11	55.02	55.27	53.19	53.21	53.27
	DT	61.11	61.22	61.29	56.28	56.32	56.38	54.21	54.22	54.28
	SVM	62.33	62.43	62.31	58.29	58.26	58.22	55.37	55.35	55.36
	MNB	64.28	64.23	64.44	60.31	60.33	60.39	57.21	57.34	57.28
	GBDT	66.43	66.32	66.56	62.46	62.56	62.55	59.39	59.36	59.33
	CNN	68.57	68.45	68.67	65.29	65.21	65.32	61.67	61.66	61.63
	LSTM	71.33	71.32	71.38	67.71	67.34	67.39	63.32	63.39	63.37
	CNN + LSTM	73.44	73.56	73.67	70.11	70.23	70.32	66.39	66.32	66.42
	EI-USD	75.46	75.77	75.67	72.44	72.38	72.39	69.22	69.19	69.38
TF-IWF	LR	60.46	60.68	60.48	57.22	57.31	57.26	55.23	55.11	55.28
	KNN	62.31	62.36	62.37	59.26	59.22	59.31	58.72	58.56	58.67
	DT	63.33	63.28	63.38	60.22	60.32	60.34	59.33	59.37	59.42
	SVM	64.45	64.34	64.28	61.25	61.24	61.36	60.34	60.33	60.36
	MNB	65.72	65.33	65.31	61.33	61.37	61.39	59.33	59.36	59.37
	GBDT	67.33	67.39	67.34	63.44	63.45	63.48	61.37	61.39	61.42
	CNN	70.67	70.49	70.61	67.38	67.39	67.41	64.27	64.28	64.33
	LSTM	73.45	73.65	73.46	70.21	70.22	70.27	67.46	67.38	67.39
	CNN + LSTM	75.28	75.31	75.37	72.37	72.38	72.42	69.66	69.67	69.71
	EI-USD	77.45	77.35	77.38	74.31	74.44	74.51	72.55	72.52	72.51
Word2Vec	LR	62.56	62.69	62.59	59.26	59.31	59.37	57.33	57.28	57.33
	KNN	64.46	64.56	64.44	61.11	61.02	61.27	60.19	60.21	60.27
	DT	65.54	65.43	65.52	63.12	63.26	63.22	62.37	62.35	62.36
	SVM	67.28	67.23	67.24	65.31	65.33	65.39	64.44	64.34	64.38
	MNB	69.43	69.28	69.22	66.46	66.56	66.55	65.39	65.36	65.47
	GBDT	72.27	72.31	72.21	67.29	67.21	67.32	65.67	65.66	65.63
	CNN	74.33	74.28	74.23	68.71	68.34	68.39	66.32	66.39	66.37
	LSTM	76.44	76.56	76.37	72.11	72.23	72.32	67.39	67.32	67.42
	CNN + LSTM	78.46	78.57	78.43	75.44	75.38	75.49	73.22	73.19	73.38
	EI-USD	79.56	79.64	79.53	76.56	76.53	76.57	74.33	74.23	74.31
GloVe	LR	65.46	65.38	65.48	63.22	63.31	63.26	60.23	60.11	60.28
	KNN	66.31	66.36	66.37	64.26	64.22	64.31	62.72	62.56	62.67
	DT	68.45	68.34	68.28	66.25	66.24	66.36	65.34	65.33	65.36
	SVM	70.72	70.37	70.31	67.33	67.37	67.39	66.33	66.36	66.37
	MNB	72.67	72.67	72.65	69.44	69.45	69.48	67.37	67.39	67.42
	GBDT	74.67	74.79	74.67	72.38	72.39	72.41	69.27	69.28	69.33
	CNN	77.45	77.65	77.46	74.21	74.22	74.27	72.46	72.38	72.39
	LSTM	80.28	80.31	80.37	76.37	76.38	76.42	74.66	74.67	74.71
	CNN + LSTM	83.55	83.35	83.38	78.31	78.44	78.51	76.55	76.52	76.51
	EI-USD	84.56	84.43	84.44	81.11	81.18	81.21	78.67	78.59	78.68
fastText	LR	67.23	67.66	67.45	64.27	64.38	64.33	62.28	62.38	62.39
	KNN	69.31	69.36	69.37	66.38	66.21	66.38	64.31	64.23	64.38
	DT	70.33	70.48	70.46	67.44	67.45	67.49	66.41	66.56	66.31
	SVM	73.11	73.21	73.39	71.29	71.31	71.33	69.41	69.32	69.19
	MNB	75.33	75.56	75.53	72.39	72.37	72.42	71.22	71.37	71.41
	GBDT	77.66	77.67	77.58	74.77	74.78	74.81	73.28	73.31	73.33
	CNN	79.77	79.67	79.63	76.33	76.28	76.41	75.33	75.38	75.37
	LSTM	83.43	83.44	83.49	82.33	82.38	82.31	81.29	81.19	81.28
	CNN + LSTM	85.56	85.51	85.53	84.39	84.42	84.45	82.56	82.49	82.51
	EI-USD	88.77	88.65	88.67	86.78	86.76	86.82	85.77	85.81	85.83
BERT	LR	70.33	70.38	70.36	67.43	67.44	67.43	64.33	64.18	64.28
	KNN	71.36	71.38	71.42	68.11	68.28	68.22	66.78	66.72	66.65
	DT	72.38	72.65	72.67	70.11	70.21	70.24	68.19	68.19	68.22
	SVM	74.45	74.66	74.76	72.29	72.31	72.33	70.11	70.19	70.31
	MNB	77.21	77.33	77.22	74.49	74.56	74.43	72.11	72.22	72.31
	GBDT	80.33	80.41	80.28	78.19	78.21	78.29	76.39	76.31	76.41
	CNN	82.78	82.79	82.67	80.78	80.56	80.61	79.11	79.29	79.39
	LSTM	85.33	85.38	85.39	84.21	84.38	84.41	82.48	82.41	82.48
	CNN + LSTM	88.21	88.19	88.32	87.89	87.67	87.78	85.45	85.43	85.44
	EI-USD	94.76	94.68	**94.78**	93.78	93.58	**93.62**	90.77	90.81	**90.83**

Note. The highest attained the results are presented in bold.

shown in Figure 6. The following subsections present in detail each of these aspects.

4.7.1. Batch size

Firstly, we are investigating the batch size impact of the proposed model (EI-USD) and other machine learning classifiers. Generally speaking, the batch size is an important parameter that impacts the learning dynamic method. In this batch size, we compared various batch size numbers from 50 to 250. The results indicate that the proposed model is constant on keep the performance with the different sizes and still the best performing classifier. Whereas,

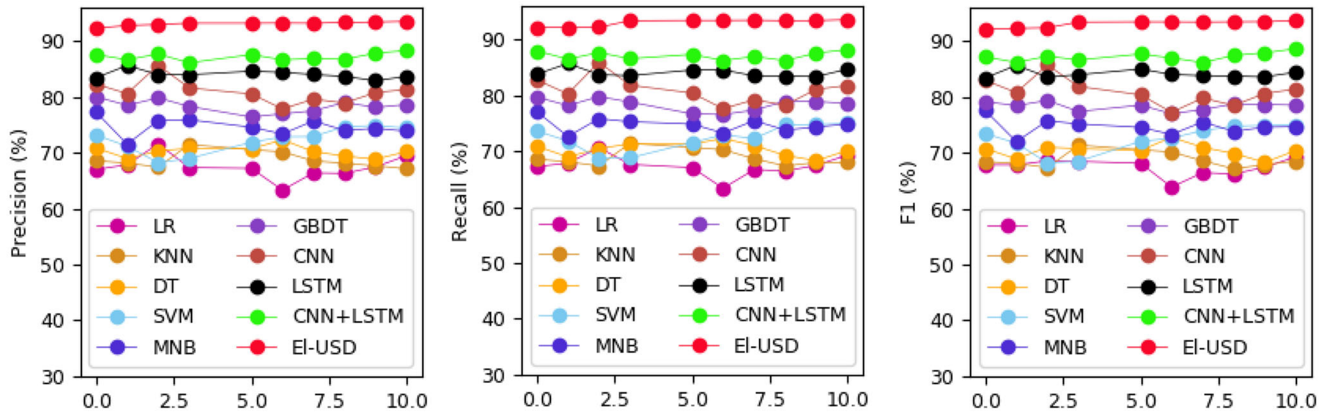

Figure 5. The results of 10-fold cross-validation of machine learning classifiers.

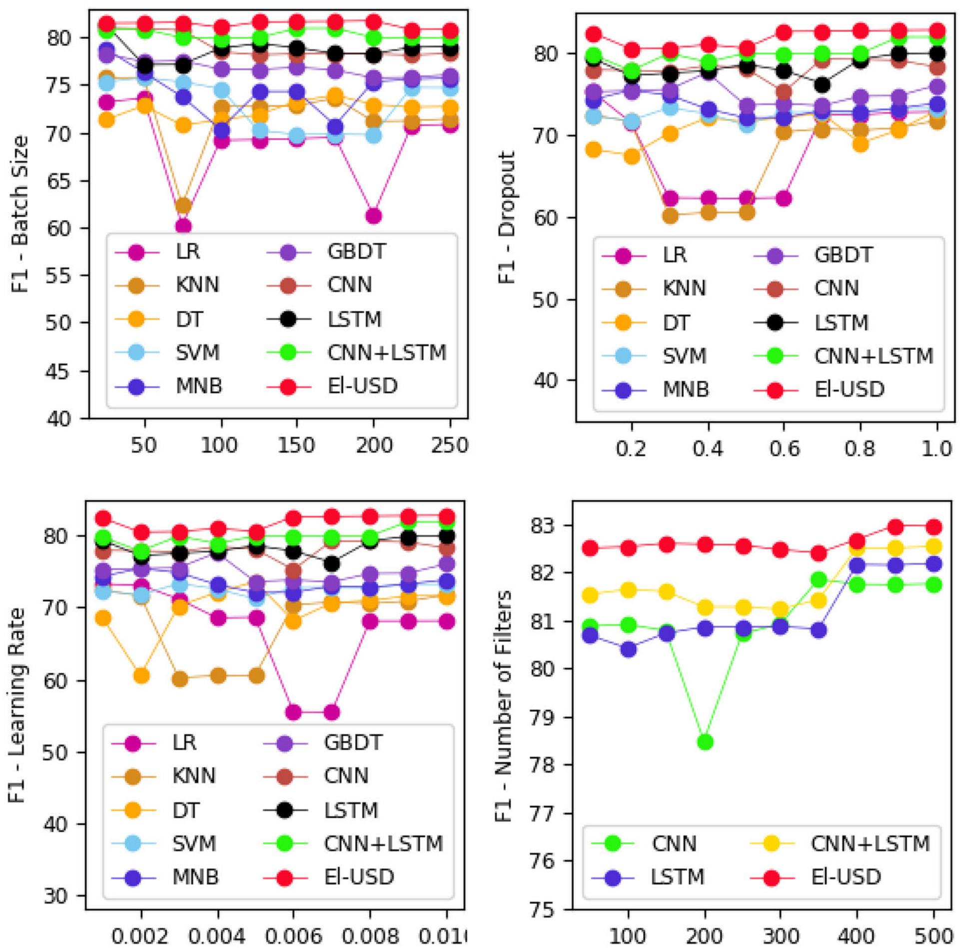

Figure 6. The machine learning classifiers' parameters impact.

the KNN and LR classifiers are highly sensitive to batch size when changing the batch size.

4.7.2. Dropout impact

To prevent avoid the over-fitting phenomenon in the machine learning classifiers, the dropout mechanism was introduced. In this experiment, we tested the machine learning classifiers with the different dropout sizes between 0.0 and 1.0. Furthermore, we employed the same value of dropout to evaluate analyzing users' satisfaction with the default dropout (i.e., 0.5). The experimental results denote that, the proposed model (El-USD) performs better than other machine learning classifiers.

4.7.3. Learning rate impact

The suitable choice of learning rate impact is important for the optimization of weight and offsets. In general, there are two situations in learning rate impact. First, if the learning rate impact is too large, then it is easy to surpass the extreme point and making the system unstable. Second, if the learning rate impact is too low, then the training time becomes longer. The proposed model achieved the highest performance results on a different learning rate configuration. We considered the range of learning rate between 0.002 and 0.01 to test the impact of learning rate. Furthermore, the experimental results indicate that KNN and LR classifiers are very sensitive to different learning rates.

4.7.4. Number of filters

The next step in this study is to investigate the number of filters impact among 50 and 500. Particularly, the results indicate that when increasing the number of filters, the proposed model performs better than other machine learning classifiers. Where, when the number 450 and 500 the proposed model performs constant on different number of filters. Furthermore, the CNN and LSTM classifier also constantly increase of performance on different number of filters.

Table 4. The analysis of ANOVA of machine learning algorithms.

F1 values of feature modelling techniques

Source	DF	Adj SS	Adj MS	*F*-value	*p*-Value	*F*-critical
Machine learning classifiers	10	6549.8	727.758	257.92	0	1.606
Error	90	253.9	2.822			
Total	99	6803.8				

Models	Total CV	Mean	*SD*	95% CI
LR	10	67.36	2.627	(66.306, 68.417)
KNN	10	68.76	2.340	(67.706, 69.817)
DT	10	70.37	2.301	(69.317, 71.428)
SVM	10	72.47	1.809	(71.415, 73.526)
MNB	10	74.66	1.501	(73.600, 75.711)
GBDT	10	78.32	1.466	(77.267, 79.378)
CNN	10	80.89	1.101	(79.839, 81.949)
LSTM	10	84.093	0.846	(83.038, 85.149)
CNN + LSTM	10	87.18	0.770	(86.125, 88.235)
El-USD	10	**93.08**	**0.591**	(92.025, 94.136)

Pooled *SD* = 1.67978.
Note. The highest attained the results are presented in bold.

4.8. Statistical analysis

To evaluate the statistical significance of the results presented in Section 4.6, we have performed a one-way analysis of variance (ANOVA) test in the statistical program, Minitab 18. Where Degrees of Freedom (DF) is defined as the quantity of information in the dataset, Sum of Squares (SS) represents the term reflecting the quantity of variation that is described by each model term, and Mean Squares (MS) defines the values calculated by dividing SS value by the corresponding degrees of freedom and Standard Deviation (SD) defines a statistic that assesses the dispersion of a dataset of relative values to the mean and is computed as the root of square of the variance. The Adjusted Sum of Squares (Adj SS) is defined as the variation measuring different components of the models. Whereas, the adjusted Mean Squares (Adj MS) is defined as the variation quantity which is explained by each term of model. *F*-value is the test statistic for identifying whether a term is associated with the response. Moreover, the probability (*p*-value) defines a decision about the statistical significance of the term model.

The goal is to find out whether there is a significant difference in results (in terms of F1) between the proposed model (El-USD) with other algorithms based on BERT algorithm. We employ ANOVA analysis in the 10-fold cross-validation results for that purpose. The test result of the one-way ANOVA analysis is shown in Table 4, we obtained an F − Value of 257.92 for F1 values machine learning classifiers. From the results, we argue that the factor (different classification techniques) leads to a significant difference (F-value > F-critical and score of p-value < α)[4] for the means of at least six feature modeling techniques. Based on the F1 metric, the El-USD based on BERT outperforms other techniques as shown in Figure 7.

4.9. Usage-based metrics results

To investigate the usability of e-learning system, usage-based metrics (i.e., completion rate, task duration, lostness, and

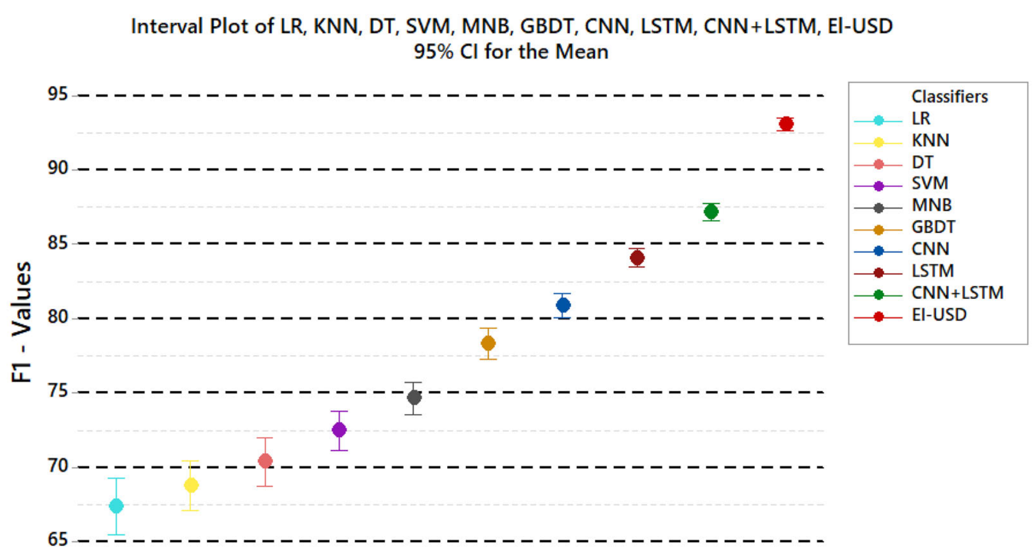

Figure 7. ANOVA analysis.

SUS) are calculated automatically based on the recorded traces of lecturers (notated as L) and students (notated as S). These metrics are calculated individually for every higher-level task for every lecturers and students *via* predefined task descriptor. Table 5 shows the summary of usage-based metrics and statistical values from all data calculated for lecturers and students. The Mean and Standard Deviation (*SD*) values are provided across all dimensions for better usability analysis. Generally, students' results are better than those of the lecturers in terms of completion rate, task duration, lostness, and SUS metrics.

Figure 8 presents the summative results achieved based on the task duration and lostness metrics for five tasks. The error bars in the plot on the lecturers' and students' data correspond to the Standard Deviation (SD) of the assessment. A considerable gap is observed between lecturers' and students' results showing that lecturers have lower values compared to the students' average values. For the case of tasks 1 up to 5, there is a variance (of 0.69 min) between lecturers and students in terms of duration. Moreover, there are different steps (0.05) between lecturers and students which is the same for all tasks. This can be the indication of deficiency in the usability of the designed e-learning system interface at this phase. An issue that needs to be addressed by the institution.

4.10. System Usability Scale (SUS) evaluation

Completion rate, task duration, and lostness metrics, respectively show to what extent a user completed a given task, time spent to complete the task, and average steps taken to complete the task. In general, based on SUS metric students have been reported to be more satisfied than lecturers as shown in Table 5. Results show that, on average lecturers have rated 77.49, whereas students have rated 83.48 on SUS. Furthermore, we use Cronbach's Alpha α motivated by Borkowska and Jach (Borkowska & Jach, 2017) to evaluate the reliability of SUS scores.

The evaluation results suggest that the scores are reliable as they have attained an α of 0.826. This is an indication of the SUS questionnaire's strong reliability as used in the lecturers' and students' satisfaction evaluations. We also note that lecturers and students are satisfied with using an e-learning system because the total number of completed activities are higher than the uncompleted ones.

Table 5. Summative evaluation for usage-based metrics per task.

Task name	Completion rate (%) L Mean	Completion rate (%) S Mean	Task duration L Mean	Task duration S Mean	Task duration *SD*	Lostness L Mean	Lostness S Mean	Lostness *SD*	SUS L Mean	SUS S Mean
Task 1	97.53	98.44	2.86	1.77	0.770	0.15	0.08	0.050	78.08	83.45
Task 2	95.65	97.49	2.92	1.97	0.670	0.17	0.14	0.025	75.28	84.31
Task 3	97.35	97.25	2.04	2.04	0.001	0.13	0.15	0.014	76.82	81.27
Task 4	95.03	97.65	3.51	2.10	0.995	0.28	0.14	0.095	78.49	83.65
Task 5	98.85	98.86	2.13	2.10	0.017	0.06	0.06	0.000	78.81	84.72
Average	96.88	97.94	2.69	2.00	0.490	0.16	0.11	0.031	77.49	83.48

4.11. The SEM evaluation

In this study, we are analyzing the convergent and discriminant validity, and reliability test scores of each variable. The convergent and discriminant validity are employed to asses a correlation value between indicator and construct scores (Gunawan et al., 2019). Reliability is a stability measure or the consistency of test scores. The reliability and consistency of the construct in this study were assessed by employing Cronbach's alpha reliability coefficient. Figure 9 shows the PLS structural model analysis.

In assessing the assessment model, Hair et al. suggested employing construct reliability, which includes Cronbach's alpha (CA), Composite Reliability (CR), and Dijkstra–Henseler rho (PA) (Hair et al., 2017). Furthermore, validity includes Discriminant ad Convergent Validity.

In Determining construct reliability, the CA value was found to be within the range of 0.831 to 0.944 as shown in Table 6. The threshold value (0.7) is lower than these figures (Almaiah & Al-Khasawneh, 2020). According to Table 6, the values of CR range from 0.880 to 0.958, which exceed the threshold value (Kline, 2015). Rather than these two values, we believe that the researches should be employed the Dijkstra–Henseler rho (ρA) reliability coefficient for evaluating and reporting the construct's reliabilities (Dijkstra &

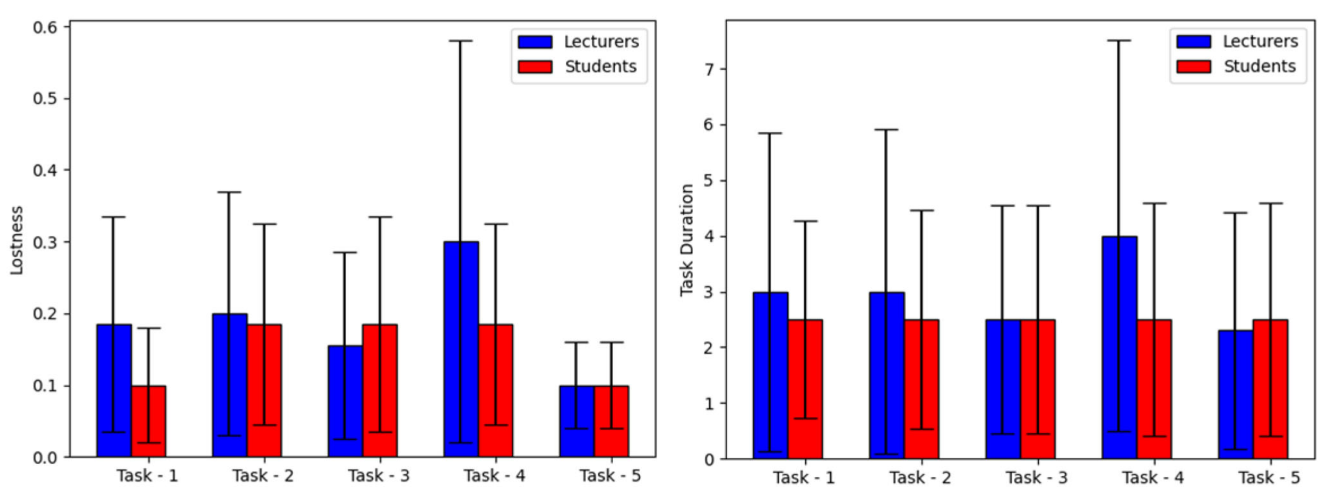

Figure 8. The summative of task duration and lostness metrics.

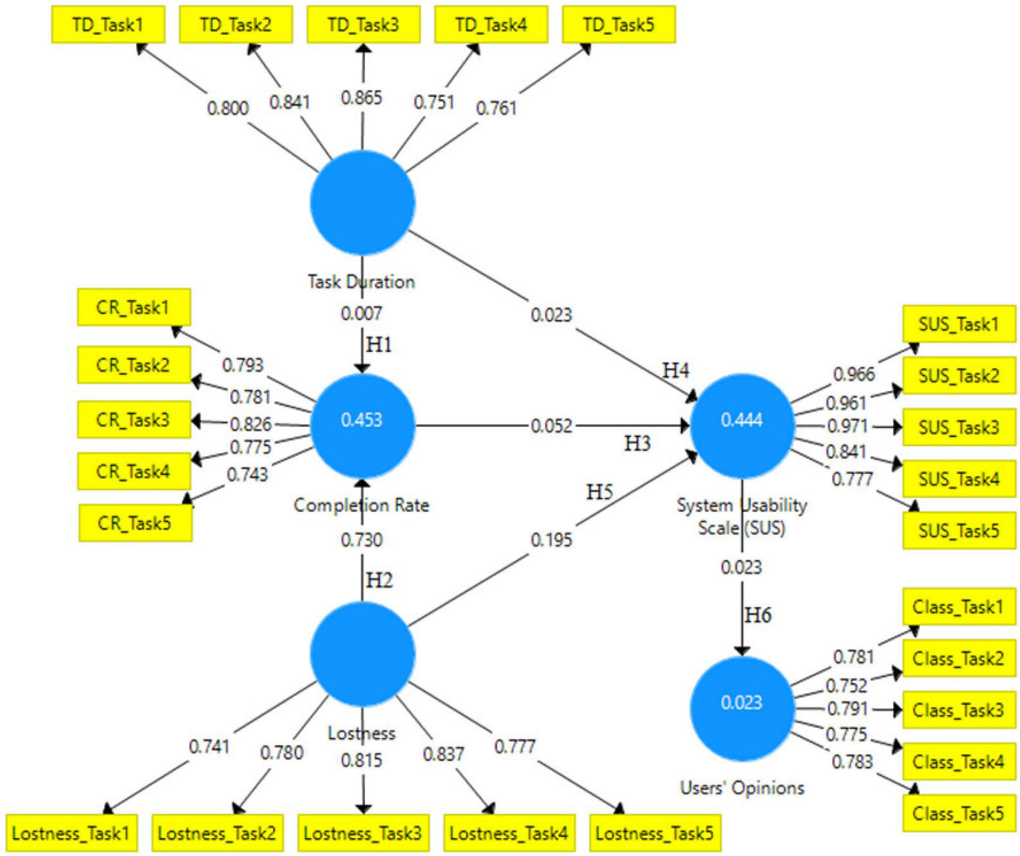

Figure 9. The structural model analysis results.

Table 6. The reliability results and analysis of convergent validity.

Latent construct	CA >0.7	ρA >0.7	CR >0.7	AVE >0.5
Completion rate	0.844	0.858	0.888	0.714
Lostness	0.831	0.838	0.880	0.796
System Usability Scale	0.944	0.957	0.958	0.822
Task duration	0.870	0.902	0.902	0.749
Users' opinions	0.848	0.855	0.884	0.703

Henseler, 2015). As with CA and CR, the coefficient of reliability ρA should be at least 0.70 (in exploratory research) and 0.80–0.90 (in advanced research stages) (Almaiah & Al-Khasawneh, 2020; Hair et al., 2011; Henseler et al., 2009).

Therefore, these values of results confirm that the constructs' reliability, and each construct was ultimately considered to be free of errors. As far as the convergent validity assessment is concerned, it is extremely important for testing the mean variance extracted (AVE) and factor loading (Hair et al., 2017). The values of AVE should be the minimum limit criteria of 0.5 (Almaiah, Al-Lozi, et al., 2021). Table 6 represents that the values of AVE ranged from 0.703 to 0.822, which exceeds the 0.5 threshold value. As a result, due to the previously mentioned explanations, it is possible that our study has convergent validity.

To measure discriminant validity, it was suggested to consider two criteria: the Heterotrait–Monotrait ratio (HTMT) and Fornell–Larcker criterion (Hair et al., 2017). Fornell et al. argued that the assessment of discriminant validity is employed to test the indicators with latent variables accuracy, which are computed by comparing the square of root value of AVE with each construct with a value of correlation between the construct and other constructs so that it can show the good values of discriminant validity that is >0.7 (Fornell et al., 1982). The findings in Table 7 suggest that the Fornell–Larker condition confirms the requirements because each AVE and its square roots exceed the value's correlation with other constructs (Fornell & Larcker, 1981).

Table 8 shows the HTMT ratio findings, which indicate that the value of each construct is lower than the 0.85 thresholds (Henseler et al., 2015). Consequently, the HTMT ratio was found to be present. With the help of these findings, the discriminant validity was calculated. According to the analysis results, there was no issue in assessing the reliability and validity of the measurement model. Consequently, the collected data were further used for evaluating the structural model.

4.12. Model fit

Urbach argued that the RMS_theta, NFI, Chi-Square, d_ULS, d_G, exact fit criteria, and standard root mean square residual (SRMR) which show the model fit in PLS-SEM are the fit measures provided by SmartPLS (Urbach Frederik, 2010). A good model fit is considered to be <0.08 SRMR values to be >0.90 NFI values (Barclay et al., 1995; Leguina, 2015). The NFI ratio deals with the Chi Square (notated as Chi^2) value in the proposed model and the null model or benchmark model (Almaiah, Al-Khasawneh, et al., 2021). The NFI is directly correlated to the parameters and

Table 7. The results of Fornell–Larcker Scale.

	Completion rate	Lostness	System Usability Scale (SUS)	Task duration	Users' opinions
Completion rate	**0.884**				
Lostness	0.672	**0.874**			
System Usability Scale (SUS)	0.657	0.631	**0.883**		
Task duration	0.631	0.653	0.684	**0.825**	
Users' opinions	0.626	0.617	0.651	0.6222	**0.877**

Note. The highest attained the results are presented in bold.

Table 8. The results of Heterotrait–Monotrait ratio (HTMT).

	Completion rate	Lostness	System Usability Scale (SUS)	Task duration	Users' opinions
Completion rate					
Lostness	0.778				
System Usability Scale (SUS)	0.605	0.698			
Task duration	0.239	0.265	0.285		
Users' opinions	0.257	0.183	0.145	0.071	

Table 9. The indicators of model fit.

	Complete model	
	Saturated model	Estimated model
SRMR	0.063	0.067
d_ULS	0.711	2.209
d_G	0.551	0.551
Chi-square	472.707	472.863
NFI	0.914	0.923
Rms theta	0.063	

Table 10. The hypotheses testing of the relationship model results.

No	Path	β	SE	t-Value	Hypothesis supported
H1	Task duration → Completion rate	0.337	0.043	2.331	Yes
H2	Lostness → Completion rate	0.411	0.071	3.452	Yes
H3	Completion rate → SUS	0.421	0.062	4.219	Yes
H4	Task duration → SUS	0.312	0.077	5.442	Yes
H5	Lostness → SUS	0.231	0.065	4.298	Yes
H6	SUS → Users' opinions	0.341	0.077	3.312	Yes

considering this, model fit indicators do not include NPI (Leguina, 2015).

The discrepancy between empirical covariance matrix and covariance matrix implied by the composite factor model is offered by the two metrics, the geodesic distance d_G, squared Euclidean distance, and d_ULS (Leguina, 2015). RMS theta can only be applied to the reflective models and helps with evaluating the degree of outer model residual correlation (Almaiah, Al-Khasawneh, et al., 2021). The PLS-SEM model will improve as the RMS theta value reaches zero, with a good fit being <0.12 and poor fits being other values (Kline, 2015). According to the suggestion of Leguina (Leguina, 2015), the relationship between each construct is evaluated by the saturated model, while the estimated model works on model structure and total effects. According to Table 9, the value of RMS_theta was 0.063. From this, it can be said that the size of the goodness-of-fit for the PLS-SEM model was appropriate for demonstrating global PLS model validity.

4.13. The testing of research-model using PLS-SEM

To determine whether the model of structural theoretical constructs are independent and thus a complete analysis of the proposed hypotheses, we exploited the Structural Equation Modelling (SEM) alongside SmartPLS with the maximum likelihood estimation (Al-Maroof et al., 2021; Al-Maroof et al., 2020). The SEM analysis results indicated that all hypotheses in the proposed model in analyzing e-learning users' satisfaction are supported as shown in Table 10.

The results indicated that Task Duration has a significantly positive correlation with Completion Rate (β-value $= 0.337$, $p < 0.001$), with this result supporting hypothesis H1. The result also found that the Completion Rate has a significantly positive relationship with System Usability Scale (SUS) (β-value $= 0.421$, $p < 0.001$), these results indicated that H3 accepted. Furthermore, H5 was also supported according to the results, which indicated that Lostness has a significant relationship with System Usability Scale (SUS) (β-value $= 0.231$, $p < 0.001$). Moreover, we also found that the System Usability Scale (SUS) has a significantly positive relationship with Users' Opinions (β-value $= 0.341$, $p < 0.001$), with these results supporting H6. Thus, the results indicated that the hypotheses H1, H2, H3, H4, H5, and H6 were also supported.

4.14. Discussion

We have compared the SEM technique and Machine Learning Algorithms (MLA). SEM is a technique based on a linear regression model used in several education domains. Siyal et al. and Akgül et al. argued that the SEM technique is more advantageous to causal the inference of relationships between independent variables and the dependent variable (Akgül & Uymaz, 2022; Siyal et al., 2021). However, these basic linear regression models may be inadequate for representing the complexity of real-world the challenges of decision making. Whereas, the machine learning algorithms can produce the reasonably advanced models of non-linear regression with higher accuracy as a supplement to linear models can be employed. From our experimental results have empirically confirmed that machine learning algorithms have advantageous predictive power but SEM's superior explanation capability. With an in-depth analysis, the prediction advantages of machine learning-algorithms can be understood in its capabilities of simultaneously capturing variables' multi-level relationships and weighting them with a large parameter space for an adaptive use. The experimental results confirm that using machine learning classifiers more accurate, effective, and can be reliably used to analyze the level of users' satisfaction in e-learning systems.

On the other hand, in this study, TAM cannot explore more deeply about the users' activities, usability testing, and opinions carried out in e-learning system. Ajibade argued that TAM model is not robust enough to explain user's behavior when the users interact in the system (Ajibade, 2019). Whereas, usage-based, SUS metrics, and machine learning classifiers can evaluate the activities carried out by the users. In addition, the existence of usage-based metrics proves that the system can be accepted based on the values carried out objectively.

The evaluation of users' satisfaction by employing machine learning classifiers offers an additional benefit of automating the task rather than relying on the manual analysis as the previous works. As it could be observed in the correlation of SUS metric and opinions-based analysis of users' satisfaction as shown in Table 11. It is worth noting that, the number of opinions that were classified as neutral (13.25%) are not considered in our analysis because they imply that there was no enough information to justify whether lecturers and students are satisfied or not. Moreover, we performed a further investigation on how reliable the opinions' sentiments (i.e., positive and negative) can be used to infer users' satisfaction. To accomplish that, we analyze users' scores on SUS and the sentiments of their opinions.

Table 11. The results of usage-based metrics, SUS metric and users' opinions.

	Lecturers				Students			
	SUS > 70		SUS < 70		SUS > 70		SUS < 70	
Task name	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Task 1	178	88	90	97	5803	1750	1773	2528
Task 2	174	111	91	121	5433	2310	2103	2708
Task 3	201	79	56	90	5156	2477	1821	3023
Task 4	220	95	47	71	5324	2590	1864	3169
Task 5	212	96	58	74	6043	2550	1636	2831

Consequently, for each task, we counted the number of positive (notated as Pos) and negative (notated as Neg) opinions under each category of SUS scores (i.e., SUS > 70 and SUS < 70). The results of this analysis are depicted in Table 11. Generally, from the table we can observe that, the number of positive opinions are higher than negative opinions for the SUS > 70, and otherwise for the SUS < 70. This suggests that satisfied users more often provide positive opinions than unsatisfied users who often give negative opinions. For instance, 66.92% of the lecturers in Task-1 who scored SUS > 70 had positive opinions, whereas the remaining 33.08% had negative opinions. A similar trend can also be observed for the same task who had SUS > 70, where 76.83% of such students had positive opinions and only 23.17% had negative opinions. To describe the users' opinions results based on five-tasks are depicted in Figure 10. Therefore, from the preceding analysis, we conclude that the sentiments of the opinions can be reliably used to analyze users' satisfaction.

Finally, we compared our proposed method with the common usage-based metrics approach. In the comparison, we examined the correlation between the results of usage-based metrics and the classification report on the number of opinions in each class. We also computed the Pearson Correlation Coefficient (PCC) between the usage-based metrics and obtained an average PCC value of $r = 0.736$. Furthermore, PCC for the usage-based, SUS metrics, and machine learning classifiers we obtained an average PCC value of $r = 0.778$. Such results lead us to the conclusion that the two approaches significantly correlate. Therefore, the findings of this study confirm that the proposed approach is effective in analyzing users' satisfaction. As it could be observed, the users in the Management *Mn* field are more satisfied with expressed positive opinions than other fields as shown in Figure 11.

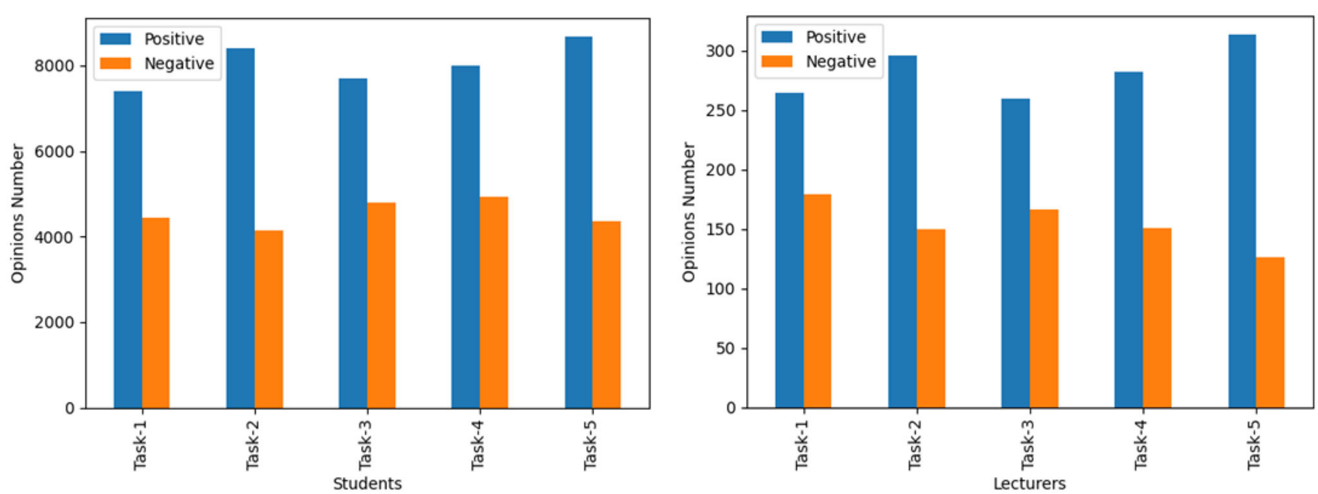

Figure 10. The users' satisfaction results for each task.

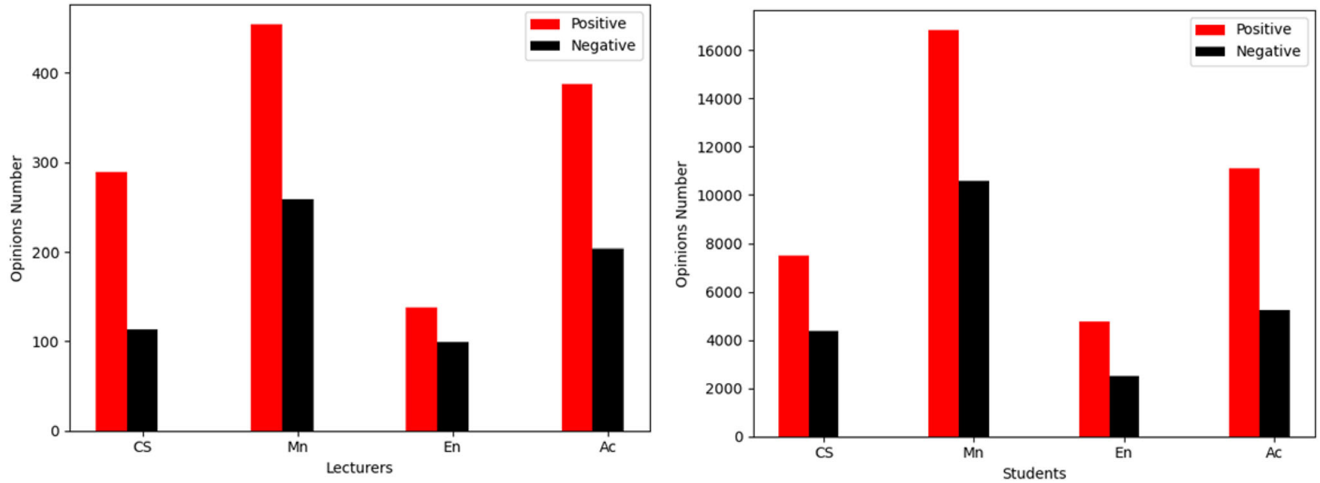

Figure 11. The users' satisfaction results are based on opinions for each field.

5. Conclusion, implication, limitation, and future work

In this study, we used traditional techniques, such as usage-based and SUS metrics to analyze and evaluate lecturers' and students' satisfaction. Usage-based and SUS metrics are used to respectively describe the activities of the users on the system, and how they perceive about the usability of the system. Moreover, we proposed a new model coined as El-USD to analyze e-learning users' satisfaction based on opinions. We have collected a text corpus containing 14,810 users' opinions for each task and classified them by using machine learning classifiers. The results indicate that the proposed model (El-USD) outperformed other classifiers with an average of 93.08% in F1 based on BERT. Furthermore, an investigation into analyzing users' satisfaction compared our proposed approach with the traditional approaches. In this case, we examined the correlation between the results of traditional approaches and that of our approach. The results indicate that there is a higher correlation between our proposed method and traditional approaches with $r = 0.778$.

Moreover, this study presents several the implications of theoretical and practical. Generally, this study findings deliver useful several suggestions to take decision makers, service providers, and designers in the universities as to how to assess and improve the e-learning system quality and understanding of multidimensional factors for effectively employing e-learning system platform. First, the designers or developers of e-learning system platforms should be focused on the factors that play a key role in improving the e-learning systems' quality, which in turn affects learning efficiency and performance of students and lecturers. Second, in this study findings show how the factors of technological pertaining to students' and lecturers' actual adopt of e-learning system platforms are significant. Therefore, the employ of e-learning system platforms should be supported by the excellent infrastructure of technology in the universities and thus this will increase the actual employ of the e-learning system platforms between students and lecturers. Third, the findings of this study can help designers or developers to develop e-learning system platforms by providing well-designed learning materials appropriate users' knowledge. Fourth, the leaders in the universities need to support distance-learning schemes in enhancing learning processes by providing the sufficient resources of financial and technological. This will lead to improvements in the system and quality of service, which positively affect students' and lecturers' behavior and thus increase the actual employ of the e-learning system.

Despite this work presenting various interesting contributions, various limitations should be covered in the future work. First, further investigations have to explore the topics of users' opinions per task for evaluating satisfaction in e-learning systems. Second, future work should explore emotions to evaluate users' satisfaction in using e-learning systems. Third, future studies can apply different sentiment word embedding techniques to analyze users' satisfaction in large training sets. Finally, there is a need to propose a novel method by combining machine learning and SEM techniques in analyzing e-learning users' satisfaction based on users' opinions, activities, and usability testing.

Notes

1. https://www.nltk.org/
2. https://scikit-learn.org/
3. https://fasilkom.mercubuana.ac.id/moeld/
4. Alpha ($\alpha = 0.005$).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Key R&D Program of China (No. 2019YFB1406302) and the National Natural Science Foundation of China (No. 61370137).

ORCID

Sulis Sandiwarno http://orcid.org/0000-0003-3370-4541

Code availability statement

The code may be accessed with authorization from the author(s) or Beijing Institute of Technology.

References

Adinolfi, P., D'Avanzo, E., Lytras, M. D., Novo-Corti, I., & Picatoste, J. (2016). Sentiment analysis to evaluate teaching performance. *International Journal of Knowledge Society Research*, *7*(4), 86–107. https://doi.org/10.4018/IJKSR.2016100108

Adler, J., & Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A*, *77A*(8), 733–742. https://doi.org/10.1002/cyto.a.20896

Ajibade, P. (2019). Technology acceptance model limitations and criticisms: Exploring the practical applications and use in technology-related studies, mixed-method, and qualitative researches. *Library Philosophy and Practice*, *9*, 1–4. http://digitalcommons.unl.edu/libphilprac/1941

Akgül, Y., & Uymaz, A. O. (2022). Facebook/Meta usage in higher education: A deep learning-based dual-stage SEM-ANN analysis. *Education and Information Technologies*, *27*(7), 9821–9855. https://doi.org/10.1007/s10639-022-11012-9

Albert, B., William, T., Tullis, T., & Tedesco, D. (2010). Beyond the usability lab: Conducting large-scale user experience studies. In *Beyond the Usability Lab*.

Al-Emran, M., & Salloum, S. A. (2017). Students' attitudes towards the use of mobile technologies in e-Evaluation. *International Journal of Interactive Mobile Technologies*, *11*(5), 95–202. https://doi.org/10.3991/ijim.v11i5.6879

Al-Fraihat, D., Joy, M., Masa'deh, R., & Sinclair, J. (2020). Evaluating E-learning systems success: An empirical study. *Computers in Human Behavior*, *102*, 67–86. https://doi.org/10.1016/j.chb.2019.08.004

Alghabban, W. G., & Hendley, R. (2022). Perceived level of usability as an evaluation metric in adaptive E-learning. *SN Computer Science*, *3*(3), 1–11. https://doi.org/10.1007/s42979-022-01138-5

AlGhannam, B. A., Albustan, S. A., Al-Hassan, A. A., & Albustan, L. A. (2018). Towards a Standard Arabic System Usability Scale: Psychometric evaluation using communication disorder app. *International Journal of Human–Computer Interaction*, *34*(9), 799–804. https://doi.org/10.1080/10447318.2017.1388099

Alkhateeb, M. A., & Abdalla, R. A. (2021). Factors influencing student satisfaction towards using learning management system Moodle. *International Journal of Information and Communication Technology Education*, *17*(1), 138–153. https://doi.org/10.4018/IJICTE.2021010109

Almaiah, M. A. (2020). Thematic analysis for classifying the main challenges and factors influencing the successful implementation of E-learning system using NVivo. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1), 142–152. https://doi.org/10.30534/ijatcse/2020/22912020

Almaiah, M. A., Alfaisal, R., Salloum, S. A., Al-Otaibi, S., Al Sawafi, O. S., Al-Maroof, R. S., Lutfi, A., Alrawad, M., Mulhem, A. A., & Awad, A. B. (2022). Determinants influencing the continuous intention to use digital technologies in Higher Education. *Electronics*, *11*(18), 2827. https://doi.org/10.3390/electronics11182827

Almaiah, M. A., Alhumaid, K., Aldhuhoori, A., Alnazzawi, N., Aburayya, A., Alfaisal, R., Salloum, S. A., Lutfi, A., Al Mulhem, A., Alkhdour, T., Awad, A. B., & Shehab, R. (2022). Factors affecting the adoption of digital information technologies in higher education: An empirical study. *Electronics*, *11*(21), 3572. https://doi.org/10.3390/electronics11213572

Almaiah, M. A., & Al-Khasawneh, A. (2020). Investigating the main determinants of mobile cloud computing adoption in university campus. *Education and Information Technologies*, *25*(4), 3087–3107. https://doi.org/10.1007/s10639-020-10120-8

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic. *Education and Information Technologies*, *25*(6), 5261–5280. https://doi.org/10.1007/s10639-020-10219-y

Almaiah, M. A., Al-Khasawneh, A., Althunibat, A., & Almomani, O. (2021). Exploring the main determinants of mobile learning application usage during Covid-19 pandemic in Jordanian Universities. In *Studies in systems, decision and control* (Vol. 348). Springer. https://doi.org/10.1007/978-3-030-67716-9_17

Almaiah, M. A., Al-Lozi, E. M., Al-Khasawneh, A., Shishakly, R., & Nachouki, M. (2021). Factors affecting students' acceptance of mobile learning application in higher education during covid-19 using ANN-SEM modelling technique. *Electronics*, *10*(24), 3121. https://doi.org/10.3390/electronics10243121

Almaiah, M. A., Al-Otaibi, S., Lutfi, A., Almomani, O., Awajan, A., Alsaaidah, A., Alrawad, M., & Awad, A. B. (2022). Employing the TAM model to investigate the readiness of M-learning system usage using SEM technique. *Electronics*, *11*(8), 1259. https://doi.org/10.3390/electronics11081259

Almaiah, M. A., Ayouni, S., Hajjej, F., Lutfi, A., Almomani, O., & Awad, A. B. (2022). Smart mobile learning success model for higher educational institutions in the context of the COVID-19 pandemic. *Electronics*, *11*(8), 1278. https://doi.org/10.3390/electronics11081278

Almaiah, M. A., Hajjej, F., Lutfi, A., Al-Khasawneh, A., Alkhdour, T., Almomani, O., & Shehab, R. (2022). A conceptual framework for determining quality requirements for mobile learning applications using Delphi method. *Electronics*, *11*(5), 788. https://doi.org/10.3390/electronics11050788

Almaiah, M. A., Hajjej, F., Lutfi, A., Al-Khasawneh, A., Shehab, R., Al-Otaibi, S., & Alrawad, M. (2022). Explaining the factors affecting students' attitudes to using online learning (Madrasati platform) during COVID-19. *Electronics*, *11*(7), 973. https://doi.org/10.3390/electronics11070973

Almaiah, M. A., Jalil, M. M. A., & Man, M. (2016). Empirical investigation to explore factors that achieve high quality of mobile learning system based on students' perspectives. *Engineering Science and Technology, An International Journal*, *19*(3), 1314–1320. https://doi.org/10.1016/j.jestch.2016.03.004

Almarashdeh, I. (2016). Sharing instructors experience of learning management system: A technology perspective of user satisfaction in distance learning course. *Computers in Human Behavior*, *63*, 249–255. https://doi.org/10.1016/j.chb.2016.05.013

Al-Maroof, R., Ayoubi, K., Alhumaid, K., Aburayya, A., Alshurideh, M., Alfaisal, R., & Salloum, S. (2021). The acceptance of social media video for knowledge acquisition, sharing and application: A comparative study among YouYube users and TikTok users' for medical purposes. *International Journal of Data and Network Science*, *5*(3), 197–214. https://doi.org/10.5267/j.ijdns.2021.6.013

Al-Maroof, R. S., Salloum, S. A., AlHamadand, A. Q. M., & Shaalan, K. (2020). Understanding an extension technology acceptance model of google translation: A multi-cultural study in United Arab Emirates. *International Journal of Interactive Mobile Technologies*, *14*(3), 157–178. https://doi.org/10.3991/ijim.v14i03.11110

Althunibat, A., Almaiah, M. A., & Altarawneh, F. (2021). Examining the factors influencing the mobile learning applications usage in higher education during the covid-19 pandemic. *Electronics*, *10*(21), 2676. https://doi.org/10.3390/electronics10212676

Ani, N. (2020). Evaluation method of mobile health apps for the elderly. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *3307*, 388–394. https://doi.org/10.32628/CSEIT206469

Ani, N., Noprisson, H., & Ali, N. M. (2019). Measuring usability and purchase intention for online travel booking: A case study. *International Review of Applied Sciences and Engineering*, *10*(2), 165–171. https://doi.org/10.1556/1848.2019.0020

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, *2*(2), 285–324.

Barrón Estrada, M. L., Zatarain Cabada, R., Oramas Bustillos, R., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*, *150*, 113265. https://doi.org/10.1016/j.eswa.2020.113265

Batbaatar, E., Li, M., & Ryu, K. H. (2019). Semantic-emotion neural network for emotion recognition from text. *IEEE Access*, *7*, 111866–111878. https://doi.org/10.1109/ACCESS.2019.2934529

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1–4). Springer.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Borkowska, A., & Jach, K. (2017). Pre-testing of Polish translation of System Usability Scale (SUS). In *Advances in intelligent systems and computing* (p. 521). Springer. https://doi.org/10.1007/978-3-319-46583-8_12

Brooke, J. (1996). SUS – A quick and dirty usability scale. In *Usability evaluation in industry*. Taylor & Francis. https://doi.org/10.1002/hbm.20701

Cohen, A., & Baruth, O. (2017). Personality, learning, and satisfaction in fully online academic courses. *Computers in Human Behavior*, *72*, 1–12. https://doi.org/10.1016/j.chb.2017.02.030

Dijkstra, T. K., & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, *81*, 10–23. https://doi.org/10.1016/j.csda.2014.07.008

Duin, A. H., & Tham, J. (2020). The current state of analytics: Implications for learning management system (LMS) use in writing pedagogy. *Computers and Composition*, *55*, 102544. https://doi.org/10.1016/j.compcom.2020.102544

Elia, G., Solazzo, G., Lorenzo, G., & Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, *92*, 589–599. https://doi.org/10.1016/j.chb.2018.04.033

Feng, S., Wang, D., Yu, G., Gao, W., & Wong, K. F. (2011). Extracting common emotions from blogs based on fine-grained sentiment clustering. *Knowledge and Information Systems*, *27*(2), 281–302. https://doi.org/10.1007/s10115-010-0325-9

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(3), 382–388. https://doi.org/10.1177/002224378101800104

Fornell, C., Tellis, G. J., & Zinkhan, G. M. (1982). Validity assessment: A structural equations approach using partial least squares. In *An assessment of marketing thought & practice*. Sage Publications.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*(2/3), 131–163. https://doi.org/10.1023/A:1007465528199

Giang, N. T. P., Dien, T. T., & Khoa, T. T. M. (2020). Sentiment analysis for university students' feedback. In *Advances in intelligent systems and computing* (p. 1130). AISC. https://doi.org/10.1007/978-3-030-39442-4_5

Giannakos, M. N., Mikalef, P., & Pappas, I. O. (2022). Systematic literature review of E-learning capabilities to enhance organizational learning. *Information Systems Frontiers*, *24*(2), 619–635. https://doi.org/10.1007/s10796-020-10097-2

Giray, G. (2021). An assessment of student satisfaction with e-learning: An empirical study with computer and software engineering undergraduate students in Turkey under pandemic conditions. *Education and Information Technologies*, *26*(6), 6651–6673. https://doi.org/10.1007/s10639-021-10454-x

Girish, V. G., Kim, M. Y., Sharma, I., & Lee, C. K. (2022). Examining the structural relationships among e-learning interactivity, uncertainty avoidance, and perceived risks of COVID-19: Applying extended technology acceptance model. *International Journal of Human–Computer Interaction*, *38*(8), 742–752. https://doi.org/10.1080/10447318.2021.1970430

Gunawan, H., Sinaga, B. L., & Wp, S. P. (2019). Assessment of the readiness of micro, small and medium enterprises in using E-money using the unified theory of acceptance and use of technology (UTAUT) method. *Procedia Computer Science*, *161*, 316–323. https://doi.org/10.1016/j.procs.2019.11.129

Guo, Y., Li, W., Jin, C., Duan, Y., & Wu, S. (2018). An integrated neural model for sentence classification. In *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*. https://doi.org/10.1109/CCDC.2018.8408230

Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems*, *117*(3), 442–458. https://doi.org/10.1108/IMDS-04-2016-0130

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, *19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior*, *61*, 463–471. https://doi.org/10.1016/j.chb.2016.03.051

Hasan, H. F., Nat, M., & Vanduhe, V. Z. (2019). Gamified collaborative environment in Moodle. *IEEE Access*, *7*, 89833–89844. https://doi.org/10.1109/ACCESS.2019.2926622

Hasan, K. S., & Ng, V. (2014). Automatic keyphrase extraction: A survey of the state of the art. In *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 – Proceedings of the Conference*. https://doi.org/10.3115/v1/p14-1119

He, C., Gu, J., Ji, Z., & Yang, X. (2018). Comparative study on the usability of navigation style in iteration process of mobile software. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 586–602), https://doi.org/10.1007/978-3-319-91803-7_44

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*, 277–319. https://doi.org/10.1108/S1474-7979(2009)0000020014

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, *145*, 103724. https://doi.org/10.1016/j.compedu.2019.103724

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*(4), e0249423. https://doi.org/10.1371/journal.pone.0249423

Hong, J. C., Tai, K. H., Hwang, M. Y., Kuo, Y. C., & Chen, J. S. (2017). Internet cognitive failure relevant to users' satisfaction with content and interface design to reflect continuance intention to use a government e-learning system. *Computers in Human Behavior*, *66*, 353–362. https://doi.org/10.1016/j.chb.2016.08.044

Ifinedo, P., Pyke, J., & Anwar, A. (2018). Business undergraduates' perceived use outcomes of Moodle in a blended learning environment: The roles of usability factors and external support. *Telematics and Informatics*, *35*(1), 93–102. https://doi.org/10.1016/j.tele.2017.10.001

Jafarian, H., Taghavi, A. H., Javaheri, A., & Rawassizadeh, R. (2021). Exploiting BERT to improve aspect-based sentiment analysis performance on Persian language. In *2021 7th International Conference on Web Research, ICWR 2021*. https://doi.org/10.1109/ICWR51868.2021.9443131

Jiang, L., Zhang, L., Yu, L., & Wang, D. (2019). Class-specific attribute weighted naive Bayes. *Pattern Recognition*, *88*, 321–330. https://doi.org/10.1016/j.patcog.2018.11.032

Kara, M., Kukul, V., & Çakır, R. (2021). Self-regulation in three types of online interaction: How does it predict online pre-service teachers' perceived learning and satisfaction? *The Asia-Pacific Education Researcher*, *30*(1), 1–10. https://doi.org/10.1007/s40299-020-00509-x

Kechaou, Z., Ben Ammar, M., & Alimi, A. M. (2011). Improving e-learning with sentiment analysis of users' opinions. In *2011 IEEE Global Engineering Education Conference, EDUCON 2011* (pp. 1032–1038). https://doi.org/10.1109/EDUCON.2011.5773275

Khan, F. H., Qamar, U., & Bashir, S. (2017). A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems*, *51*(3), 851–872. https://doi.org/10.1007/s10115-016-0993-1

Kim, H. j., Kim, J., Kim, J., & Lim, P. (2018). Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning. *Neurocomputing*, *315*, 128–134. https://doi.org/10.1016/j.neucom.2018.07.002

Kline, R. B. (2015). Principles and practice of structural equation modeling. In *Methodology in the social sciences*. Guilford Publications.

Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, *40*(10), 4065–4074. https://doi.org/10.1016/j.eswa.2013.01.001

Kornpitack, P., & Sawmong, S. (2022). Empirical analysis of factors influencing student satisfaction with online learning systems during the COVID-19 pandemic in Thailand. *Heliyon*, *8*(3), e09183. https://doi.org/10.1016/j.heliyon.2022.e09183

Kurucay, M., & Inan, F. A. (2017). Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course. *Computers & Education*, *115*, 20–37. https://doi.org/10.1016/j.compedu.2017.06.010

Landrum, B., Bannister, J., Garza, G., & Rhame, S. (2021). A class of one: Students' satisfaction with online learning. *Journal of Education for Business*, *96*(2), 82–88. https://doi.org/10.1080/08832323.2020.1757592

Lauscher, A., Ponzeo, S. P., Glavaš, G., & Eckert, K. (2017). Investigating convolutional networks and domain-specific embeddings for semantic classification of citations. In *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3127526.3127531

Leguina, A. (2015). A primer on partial least squares structural equation modeling (PLS-SEM). *International Journal of Research & Method in Education*, *38*(2), 220–221. https://doi.org/10.1080/1743727X.2015.1005806

Li, M., Chen, L., Zhao, J., & Li, Q. (2021). Sentiment analysis of Chinese stock reviews based on BERT model. *Applied Intelligence*, *51*(7), 5016–5024. https://doi.org/10.1007/s10489-020-02101-8

Lin, S., Salazar, T. R., & Wu, S. (2019). Impact of academic experience and school climate of diversity on student satisfaction. *Learning Environments Research*, *22*(1), 25–41. https://doi.org/10.1007/s10984-018-9265-1

Lin, Q., Zhu, Y., Zhang, S., Shi, P., Guo, Q., & Niu, Z. (2019). Lexical based automated teaching evaluation via students' short reviews. *Computer Applications in Engineering Education*, *27*(1), 194–205. https://doi.org/10.1002/cae.22068

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of natural language processing* (pp. 1–38). Oxfordshire. https://doi.org/10.1145/1772690.1772756

Liu, W., Liu, P., Yang, Y., Gao, Y., & Yi, J. (2017). An attention-based syntax-tree and tree-LSTM model for sentence summarization. *International Journal of Performability Engineering*, *13*(5), 775–782. https://doi.org/10.23940/ijpe.17.05.p20.775782

Liu, S., & Yang, G. (2020). Negative sentiment analysis of MOOC comments based on machine learning. In *Lecture Notes in Electrical Engineering* (p. 600). https://doi.org/10.1007/978-981-15-1864-5_77

Lutfi, A. (2022). Factors influencing the continuance intention to use accounting information system in Jordanian SMEs from the perspectives of UTAUT: Top management support and self-efficacy as predictor factors. *Economies*, *10*(4), 75. https://doi.org/10.3390/economies10040075

Madani, Y., Erritali, M., & Bengourram, J. (2019). Sentiment analysis using semantic similarity and Hadoop MapReduce. *Knowledge and Information Systems*, *59*(2), 413–436. https://doi.org/10.1007/s10115-018-1212-z

Maitra, S., Madan, S., Kandwal, R., & Mahajan, P. (2018). Mining authentic student feedback for faculty using Naïve Bayes classifier. *Procedia Computer Science*, *132*, 1171–1183. https://doi.org/10.1016/j.procs.2018.05.032

Mikolov, T. (2015). Efficient estimation of word representations in vector space Tomas. In *IJCAI International Joint Conference on Artificial Intelligence*. https://doi.org/10.1162/153244303322533223

Modi, N., & Chaubey, P. (2021). Learning management system in secondary education. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *18*(7), 2733–2740.

Nguyen, P. X. V., Hong, T. T. T., Nguyen, K. V., & Nguyen, N. L. T. (2019). Deep learning versus traditional classifiers on Vietnamese students' feedback corpus. In *NICS 2018 – Proceedings of 2018 5th NAFOSTED Conference on Information and Computer Science*. https://doi.org/10.1109/NICS.2018.8606837

Niu, X., Hou, Y., & Wang, P. (2017). Bi-directional LSTM with quantum attention mechanism for sentence modeling. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (p. 10635). LNCS. https://doi.org/10.1007/978-3-319-70096-0_19

Onan, A. (2020). Mining opinions from instructor evaluation reviews: A deep learning approach. *Computer Applications in Engineering Education*, *28*(1), 117–138. https://doi.org/10.1002/cae.22179

Onan, A., Korukoğlu, S., & Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Systems with Applications*, *62*, 1–16. https://doi.org/10.1016/j.eswa.2016.06.005

Oramas Bustillos, R., Zatarain Cabada, R., Barrón Estrada, M. L., & Hernández Pérez, Y. (2019). Opinion mining and emotion recognition in an intelligent learning environment. *Computer Applications in Engineering Education*, *27*(1), 90–101. https://doi.org/10.1002/cae.22059

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review*, *119*, 105535. https://doi.org/10.1016/j.childyouth.2020.105535

Patwa, P., Aguilar, G., Kar, S., Pandey, S., Srinivas, P. Y. K. L., Gambäck, B., Chakraborty, T., Solorio, T., & Das, A. (2020). SemEval-2020 Task 9: Overview of sentiment analysis of code-mixed tweets. In *14th International Workshops on Semantic Evaluation, SemEval 2020 – Co-located 28th International Conference on Computational Linguistics, COLING 2020, Proceedings*. https://doi.org/10.18653/v1/2020.semeval-1.100

Pennington, J., Socher, R., Manning, C. D. (2014). GloVe: Global vectors for word representation. In *EMNLP 2014 – 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. https://doi.org/10.3115/v1/d14-1162

Periaiya, S., & Nandukrishna, A. T. (2023). What drives user stickiness and satisfaction in OTT video streaming platforms? A mixed-method exploration. *International Journal of Human–Computer Interaction*, *39*(39), 1–17. https://doi.org/10.1080/10447318.2022.2160224

Perišić, J., Milovanović, M., & Kazi, Z. (2018). A semantic approach to enhance Moodle with personalization. *Computer Applications in Engineering Education*, *26*(4), 884–901. https://doi.org/10.1002/cae.21929

Pinho, C., Franco, M., & Mendes, L. (2021). Application of innovation diffusion theory to the E-learning process: Higher education context. *Education and Information Technologies*, *26*(1), 421–440. https://doi.org/10.1007/s10639-020-10269-2

Pota, M., Ventura, M., Catelli, R., & Esposito, M. (2020). An effective BERT-based pipeline for twitter sentiment analysis: A case study in Italian. *Sensors*, *21*(1), 133. https://doi.org/10.3390/s21010133

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*(1), 173–191. https://doi.org/10.1007/s10758-020-09446-6

Qi, C., & Liu, S. (2021). Evaluating on-line courses via reviews mining. *IEEE Access*, *9*, 35439–35451. https://doi.org/10.1109/ACCESS.2021.3062052

Revythi, A., & Tselios, N. (2019). Extension of technology acceptance model by using system usability scale to assess behavioral intention to use e-learning. *Education and Information Technologies*, *24*(4), 2341–2355. https://doi.org/10.1007/s10639-019-09869-4

Sabeh, H. N., Husin, M. H., Kee, D. M. H., Baharudin, A. S., & Abdullah, R. (2021). A systematic review of the DeLone and McLean model of information systems success in an e-learning context (2010–2020). *IEEE Access*, *9*, 81210–81235. https://doi.org/10.1109/ACCESS.2021.3084815

Salton, G. (1991). Developments in automatic text retrieval. *Science*, *253*(5023), 974–980. https://doi.org/10.1126/science.253.5023.974

Shahzad, A., Hassan, R., Aremu, A. Y., Hussain, A., & Lodhi, R. N. (2021). Effects of COVID-19 in E-learning on higher education institution students: The group comparison between male and female. *Quality & Quantity*, *55*(3), 805–826. https://doi.org/10.1007/s11135-020-01028-z

Simões, A. P., & de Moraes, A. (2012). The ergonomic evaluation of a virtual learning environment usability. *Work*, *41*(1), 1140–1144. https://doi.org/10.3233/WOR-2012-0293-1140

Siyal, A. W., Chen, H., Chen, G., Memon, M. M., & Binte, Z. (2021). Structural equation modeling and artificial neural networks approach to predict continued use of mobile taxi booking apps: The mediating role of hedonic motivation. *Data Technologies and Applications*, *55*(3), 372–399. https://doi.org/10.1108/DTA-03-2020-0066

Skrbinjek, V., & Dermol, V. (2019). Predicting students' satisfaction using a decision tree. *Tertiary Education and Management*, *25*(2), 101–113. https://doi.org/10.1007/s11233-018-09018-5

Smith, P. A. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers*, *8*(4), 365–381. https://doi.org/10.1016/S0953-5438(97)83779-4

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Tarus, J. K., Niu, Z., & Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing*, *22*(8), 2449–2461. https://doi.org/10.1007/s00500-017-2720-6

Tullis, T., & Albert, B. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics* (2nd ed.). Morgan Kaufmann. https://doi.org/10.1016/C2011-0-00016-9

Umer, Q., Liu, H., & Sultan, Y. (2018). Emotion based automated priority prediction for bug reports. *IEEE Access*, *6*, 35743–35752. https://doi.org/10.1109/ACCESS.2018.2850910

Urbach Frederik, N. A. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information Technology Theory and Application (JITTA)*, *11*(2), 5–40.

Vargas-Calderón, V., Flórez, J. S., Ardila, L. F., Parra-A, N., Camargo, J. E., & Vargas, N. (2020). Learning from students' perception on professors through opinion mining. In *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-030-61702-8_23

Vlachogianni, P., & Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education*, *54*(3), 392–409. https://doi.org/10.1080/15391523.2020.1867938

Wan, S., & Niu, Z. (2020). A hybrid e-learning recommendation approach based on learners' influence propagation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(5), 827–840. https://doi.org/10.1109/TKDE.2019.2895033

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. *IEEE Access*, *8*, 138162–138169. https://doi.org/10.1109/ACCESS.2020.3012595

Wang, X., & Zhou, R. (2022). Impacts of user expectation and disconfirmation on satisfaction and behavior intention: The moderating effect of expectation levels. *International Journal of Human–Computer Interaction*, 1–14. https://doi.org/10.1080/10447318.2022.2095479

Wang, C., Zhang, M., Ma, S., & Ru, L. (2008). Automatic online news issue construction in web environment. In *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*. https://doi.org/10.1145/1367497.1367560

Wei, H. C., & Chou, C. (2020). Online learning performance and satisfaction: Do perceptions and readiness matter? *Distance Education*, *41*(1), 48–69. https://doi.org/10.1080/01587919.2020.1724768

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, *53*(6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5

Yawson, D. E., & Yamoah, F. A. (2020). Understanding satisfaction essentials of E-learning in higher education: A multi-generational cohort perspective. *Heliyon*, *6*(11), e05519. https://doi.org/10.1016/j.heliyon.2020.e05519

Yilmaz, R. (2017). Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom. *Computers in Human Behavior*, *70*(2), 251–260. https://doi.org/10.1016/j.chb.2016.12.085

Yuen, A. H. K., Cheng, M., & Chan, F. H. F. (2019). Student satisfaction with learning management systems: A growth model of belief and use. *British Journal of Educational Technology*, *50*(5), 2520–2535. https://doi.org/10.1111/bjet.12830

Zabolotniaia, M., Cheng, Z., Dorozhkin, E. M., & Lyzhin, A. I. (2020). Use of the LMS Moodle for an effective implementation of an innovative policy in higher educational institutions. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(13), 172. https://doi.org/10.3991/ijet.v15i13.14945

Zhang, C., Lin, D., Cao, D., & Li, S. (2021). Grammar guided embedding based Chinese long text sentiment classification. *Concurrency and Computation: Practice and Experience*, *33*(21), 1–13. https://doi.org/10.1002/cpe.6439

Zhu, Y., Lin, Q., Lu, H., Shi, K., Liu, D., Chambua, J., Wan, S., & Niu, Z. (2023). Recommending learning objects through attentive heterogeneous graph convolution and operation-aware neural network. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 4178–4189. https://doi.org/10.1109/TKDE.2021.3125424

About the authors

Sulis Sandiwarno is currently pursuing the PhD degree with the School of Computer Science and Technology, Beijing Institute of Technology, China. His research interests include analysis of information system, e-learning system techniques, data mining, opinion mining, and computer programing.

Zhendong Niu is currently a Professor of the School of Computer Science and Technology, Beijing Institute of Technology, China. His research areas include digital libraries, e-learning techniques, information retrieval, and recommender systems. He serves as an editorial board member for international journal of learning technology.

Ally S. Nyamawe received the PhD degree in computer science from the Beijing Institute of Technology, China, in 2020. He is currently a Lecturer with the Department of Computer Science, University of Dodoma. His research interests include data mining and computer programming.