



Assessment of the digital competencies of university instructors through use of the machine learning method

Tzu-Chi Yang¹ 

Received: 19 November 2021 / Accepted: 9 January 2023 / Published online: 18 January 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

The explosion of COVID-19 has brought new challenges to the education industry, especially higher education. Digital competency is becoming an essential competency for higher education instructors, and how to assess instructors' digital competency is attracting increasing attention in higher education. However, most studies have used self-report questionnaires or manual reviews to assess digital competencies, which are time-consuming and potentially biased, and there is a current need for valid and effective assessment methods. To address this issue, this study uses machine learning to analyze syllabi to assess the extent to which university instructors have incorporated digital competency into their courses. The results show that not only is the proposed method feasible, but the results of the assessment using machine learning are highly consistent with those of the human assessment. This approach contributes to the assessment of digital competency in higher education institutions and provides evidence that can be used as a reference for future research on the development of digital competency in higher education institutions.

Keywords Digital competence · Syllabus analysis · Higher education · Machine learning

Introduction

During the COVID-19 pandemic, ordinary modes of teaching and learning activities were suspended. Many institutes of higher education and other organizations changed their teaching approaches and sought to provide a convenient, safe, and flexible educational environment for students. As a result of this sudden change, a new level of interest in distance learning and digital competence appeared in response to the need to deal with the challenges of online learning (Kawasaki et al. 2021; Saide

✉ Tzu-Chi Yang
tcyang.academic@gmail.com

¹ Institute of Education, National Yang Ming Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan, ROC

and Sheng 2021). Many educators reported deficiencies in their digital competence, which led to higher workloads and negative emotions during the vigilance period (Väättäjä and Ruokamo 2021). The development of digital competence development and the distance learning experiences of teachers and students require continuous attention, including support for online participation, the operation of digital tools, the use of a learning management system (LMS), and the retrieval and reading of online resources (Heidari et al. 2021), as the current state is unlikely to end any time soon and may even become a cycle (Manca and Delfino 2021). Therefore, instructors, students, and educational institutions need to be sufficiently digitally competent so that they can adapt to rapid changes in teaching methods (Blau et al. 2020).

Unlike K-12 schools, universities offer a high degree of flexibility and heterogeneous curriculums, and they provide professional skills development in various domains. For example, some university courses emphasize FTF (face-to-face) interactions or operations; some require students to use ICT technology or apply digital tools to complete learning tasks; and some require students to engage in online learning activities, in some cases even allowing students to take complete courses from home via the internet (Chiu et al. 2021). Many university faculty members and students lack sufficient digital competence to deal with sudden changes in learning delivery, but few universities have sought to properly prepare their faculty and students to use online learning technologies (Cabero-Almenara et al. 2021; Weber et al. 2018). Scholars have suggested that universities must consider the state of knowledge in faculty and students regarding educational ICT technologies and digital tools to meet the needs they encounter as the educational setting changes (Monroy García et al. 2020). Assessing whether university faculty and students have sufficient digital competence to cope with future changes is an important issue at this stage (Starkey 2020).

Digital competency has been regarded as an important competency for teachers and students in the digital era; universities are the essential stage of the development of digital competency, and university instructors are the keys to digital competency development by students. Therefore, assessing and understanding the digital competencies of university instructors is an important issue in the field of higher education. Some researchers (Domínguez-Lloria et al. 2021) have used collaborative platforms and questionnaires to assess students' competence with digital tools. Some researchers (Sánchez-Cruzado et al. 2021) used questionnaires to establish whether university instructors were sufficiently digitally competent to cope with the emergency educational model during the COVID-19 pandemic. Because recent studies suggest that university instructors are responding to help future professionals adapt to industrial and job market changes in the digital age, the assessment of instructors' digital competencies requires more research attention (Antón-Sancho et al. 2021; Sánchez-Cruzado et al. 2021). Specifically, the literature using quantitative analysis is limited by the breadth of the questionnaire items, and the use of manual analysis methods is limited by manpower and time, highlighting the current need for more extensive and effective digital competency assessment solutions (Sillat et al. 2021; Zhou 2021; Mattar et al. 2022).

To address this issue, scholars have argued that a syllabus constitutes a contract between instructors and students that describes not only the domain subject,

pedagogy, and resources involved in the course but also the skills students will develop and the tasks that students are expected to accomplish during the course (Parkes and Harris 2002). Hence, the analysis of syllabi can be used to assess the digital competence of instructors and students at universities (Schina et al. 2020). However, syllabus analysis is more time-consuming and labor-intensive than questionnaires and cannot easily be performed (Bensen and Silman 2012). We note that machine learning (ML) technology, which enables computer programs to mimic human recognition and classification of text, is considered to be an excellent solution for analyzing text by automating the task of analyzing text with minimal human intervention (Kadhim 2019; Iatrellis et al. 2020; Golowko 2021). In addition, machine learning based on a data-driven approach will produce more objective results to improve the limitations of traditional analysis (e.g., statistical methods) (Yakubu and Abubakar 2021; Barthakur et al. 2022). However, the feasibility and reliability of digital competency assessment using ML remain unanswered. In this vein, we would like to confirm how our approach contributes to the assessment of digital competency by answering the following two research questions (RQ). RQ1: How well can ML evaluate the level of digital competence applied in a course from its syllabus? RQ2: How well can ML classify the levels of digital competence present in courses relative to human evaluation? These answers will help to understand digital competencies in universities and give universities the opportunity to respond accordingly and prepare for the development of quality digital competency formation for higher education.

Literature review

Digital competence in higher education

Digital competence is the set of abilities necessary to use technology to optimize daily life (Ferrari 2013). The European Commission considers digital competence to be a key life skill and has developed the European Digital Competence Framework (DIGCOMP) as a reference framework for introducing and demonstrating digital competence. DIGCOMP identifies the key components of digital competence in the following five areas: (1) information processing, (2) communication and collaboration, (3) digital content creation, (4) security, and (5) problem solving (Carretero et al. 2017). Some studies have used digital literacy to describe these competences. Although digital competence and digital literacy are not identical, a growing number of scholars have argued that the distinction between them has been blurred by overlapping definitions and translations (Madsen et al. 2018). It is generally accepted that digital competence is a skill related to digital literacy, media literacy, ICT literacy, information literacy, and internet literacy (Esteve-Mon et al. 2019). For example, in educational settings, students use digital tools to produce and share information, which demonstrates their digital competence. To provide a consistent description, we use the term digital competence in this study. During the COVID-19 pandemic, a focus on the evaluation and development of digital competence in higher education has reached an unprecedented level (Pinto et al. 2020).

Digital competence is an important competency for current learning and future employment. Universities, as cultivators of expertise in a variety of fields, are key to providing students with a quality education (Olszewski and Crompton 2020). Understanding the current digital competency offered in universities will contribute to both academia and industry. With regard to the former, by obtaining a complete picture of the current digital technologies used by university faculty and students in their courses, universities will have the opportunity to make further recommendations and invest in research topics to address the digital divide. Regarding the latter, the use of digital technologies is inevitable in the era of Industry 4.0, and the digital competency of the future industrial workforce depends on the planned development (Bartolomé et al. 2022). As a bridge between learning and employment, universities have the opportunity to prepare students for the job market of the future and to contribute to the development of a digital industrial environment when we can effectively understand the digital competencies offered in university courses (e.g., by examining the extent to which the use of digital technologies in curricula in various fields is key to meeting current industry needs).

Digital competence evaluation and syllabus analysis

Although research on digital competence in higher education is accumulating, the digital competence required of university faculty and students in their respective areas is often lacking attention (Vorobel et al. 2021). Recent studies of digital competence in higher education have shown that the evaluation and investigation of digital competence in higher education generally use traditional methods and are still in their infancy (Zhao et al. 2021). To cope with this issue, researchers have suggested that educators design learning activities in accordance with their own digital competence and their perception of their learners' competence as well as in relation to the development of needed skills (Alarcón et al. 2020). If an instructor integrates digital competence into a course, analyzing the syllabus can help understand the instructor's level of digital competence (Guillén-Gámez and Mayorga-Fernández 2020). Based on the digital competency framework (Mattar et al. 2022) and the perspective of educational practice (Gower et al. 1983; Richardson 1990), it is reasonable to assess the digital competency levels of teachers by analyzing their syllabi, and evaluating digital competence by identifying instructional media/tools and learning activities in a course related to the skill has been considered reasonable (Lucas et al. 2021). For example, instructors may use software for direct instruction in the classroom, conduct computer-assisted instruction, implement educational games, use digital media for communication, or operate LMSs/platforms. All of these practices are evidence of instructors' digital competence levels (König et al. 2020). Some recent studies have investigated digital competence by analyzing syllabi. For example, Dubicki (Dubicki 2019) analyzed 180 syllabi to identify the outcomes of training digital competence development at a university. Beuoy and Boss (2019) analyzed syllabi to identify opportunities to support digital competence instruction and develop strategic pedagogy. Their results indicate that analyzing syllabi can effectively produce a big-picture understanding of the development of digital competence

in education. Other studies have found that the evaluation of syllabi can produce evidence for identifying the contribution of certain courses to enhancing digital competence (Boss and Drabinski 2014). In other words, analyzing syllabi to evaluate the digital competence of university faculty is a reliable and valid solution.

Method

This study uses machine learning to analyze syllabi and evaluate digital competency. The data for the study were collected by a web crawler and were preprocessed. Furthermore, we clarify the criteria for assessing digital competency and the performance levels of ML models. These details are described in the subsequent subsections.

Elaborate criteria for the assessment of digital competence

The digital competence framework DIGCOMP has been proposed to assess digital competence (Carretero et al. 2017). DIGCOMP is a critical document for assessing digital competence, and it has been adopted in many contexts (Hernández-Martín et al. 2021). Retrieving and processing information, communication, and creating and managing learning content have been proposed as categories of digital competence related to education (Ferrari 2013). However, security and problem solving are skills usually developed or shaped by long-term interaction with the digital environment (Caena and Redecker 2019). In this study, therefore, we focused on the area of *information and data literacy, communication and collaboration, and digital content creation* and selected assessment criteria using the DIGCOMP framework. Table 1 presents the assessments we used to evaluate whether the syllabus covers the development of digital competence.

Data collection and labeling

Using web crawling, we collected 1200 syllabi at random from one university. To avoid excessive variation in the textual information, we ignored syllabi of less than 30 words or three sentences. We classified the syllabi into four levels of digital competence, namely, High (H), Moderate (M), Low (L), or Not currently integrated (N). If the syllabus indicated what skills the instructor would use or whether the students were required to accomplish an activity described in the area of information and data literacy and only fit into this area, the digital competence level (DCPL) of the course was set to low (L). If the syllabus mentioned the instructors or students performing activities in an area of communication and collaboration that might also be related to information processing, we considered the syllabus to show a moderate DCPL (M). Similarly, if at least one digital content creation activity was described in the syllabus and both low and moderate levels of digital competence were implied, the syllabus was labeled having a high DCPL (H). Conversely, if no digital competence-related description appeared in the syllabus, the course was labeled as not yet

Table 1 Assessment of digital competence in an integrated syllabus from DIGCOMP

Dimensions	Description	Example of activities
Information and data literacy		
Browsing, searching, and filtering information	To search for and access online information, find relevant information, efficiently select online resources, and navigate between online sources	To use a search engine to find details about a specific topic or content in a course
Evaluating information	To gather, process, understand, and critically evaluate information	To identify a range of different sources about a topic and judge their value or validity
Storing and retrieving information	To manipulate and store information and content for easier retrieval, to organize information and data	To save, store, or tag files, content and information manage the information in course
Communication and collaboration		
Interacting through technology	To interact with a variety of digital devices and applications, understand appropriate ways of communicating through digital means, refer to different communication formats, and adapt communication modes and strategies to the specific audience	To use several digital tools to interact with others, such as VoIP, chat, email, and social media
Sharing information and content	To share with others the content of information found, to act as an intermediary, be proactive in the spreading content, and know about citation practices	To use a social networking site to ask partners to review an assignment after completing the assignment
Engaging in online citizenship	To participate in society through online engagement and seek opportunities for self-development and empowerment in using technologies and digital environments	To use some basic features of online services, such as eGovernment services, an e-hospital or medical center, or an e-bank
Collaborating through digital channels	To use technology and media in team work or collaborative co-construction and co-creation of resources, knowledge, and content	To create and discuss outputs in collaboration with others using simple digital tools (e.g., google doc)
Digital content creation		
Developing content	To create content in different formats including multi-media, edit and improve content, and produce creative expression through digital media and technologies	To present ideas to others in the class and use presentation software, images, video, and creatively
Integrating and re-elaborating	To modify, refine, and mash-up existing resources to create new, original, and relevant content and knowledge	To edit, refine, and modify the content produced by themselves or others

Table 1 (continued)

Dimensions	Description	Example of activities
Programming	To apply settings, program modification, program applications, software, and devices, understand the principles of programming, and understand what is behind a program	To modify software and applications or write program source code

integrating digital competence (N). The labeling process was conducted by two students with master's degrees in information education from a national university, and the decisions were verified by a professor with a background in information education who had participated in a campus information literacy assessment program at the university. Sample syllabi are shown in Table 2.

Data preprocessing

We removed noise from the raw data, including specs, punctuation marks, numbers, and non-English/Chinese characters, before feature extraction and classification were performed. All words were converted to lowercase after tokenization, and stop words, such as “the,” “a,” “an,” and “in,” were removed.

Feature extraction

Term frequency-inverse document frequency (TF-IDF) is a very common method for extracting features for text classification using ML. In general, TF-IDF can provide features more accurately than other algorithms, simplifying and streamlining text feature extraction (Al-Rimy et al. 2020). TF-IDF determines the importance of keywords in a document set using a weighting mechanism, whereby TF represents the frequency of a keyword in a document, which represents its importance in the document in which it appears, and IDF is the prevalence of a term across documents. The most representative terms in a particular document can be drawn from a large text set by considering both TF and IDF. This feature-extraction algorithm is suitable for classifying syllabi.

ML model building

Four common ML classifiers were used in this study, namely, support vector machine (SVM), logistic regression (logit), K-nearest neighbor (KNN), and naïve Bayes (NB). SVM functions by finding a maximal margin hyperplane such that data on one side of the hyperplane can be separated from those on its other side. Taking the high dimensionalities of the text features into account, SVM is capable of using nonlinear kernel/radial functions for classification and has been suggested for use in text classification (Joachims 1998). Logit classifiers use logical functions to model the relationships between features and specific outputs, although many complex extensions exist. Logit is also considered effective for text classification (Alsmadi and Hoon 2019). KNN uses all available data to classify cases based on similarity within a dataset. KNN is considered adequate for text classification (Mowafy et al. 2018). NB is a simple probabilistic classifier based on Bayes' theorem and is often used as a baseline for text classification (Xu 2018). These classification models were adopted in this study.

We used both the test set (20% of the dataset) and a tenfold cross-validation method to evaluate the effectiveness of the ML models. Within the tenfold cross-validation, we randomly divided the training dataset into ten equally sized subsets.

Table 2 Sample labeled syllabi

Course name/(DCPL)	Part of the syllabus	Example of DCPL descriptions
Life and biomedicine/(N)	<p>The topics include the introduction of life forms, genes and biomolecules, technology in generation of life, modification of biological molecules, diseases, and the conservation of life</p>	Not currently integrated
Advanced English writing/(L)	<p>This course aims to develop students' ability to write academic essays of one to several paragraphs, including introduction, body, and conclusion. Using process writing techniques, #IL1 {students will learn to generate ideas, using library/internet resources}, #IL1 {students will learn to generate ideas, using library/internet resources}, outlining, drafting, and revising. In addition, a variety of sentence structures that frequently used in academic English genre will be highlighted in weekly mini grammar lessons. Upon successful completion of this course, students will demonstrate competency in: (a) reading college level academic texts</p>	Information and data literacy {#IL1}
Computer-mediated communication/(M)	<p>This course #IL1 {provides students with knowledge of theories and methodology regarding the field of computer-mediated communication}. The three major research areas covered are: interpersonal communication, group communication, and organizational communication, #CO1 {Topics concerning information and communication technologies (ICTs)} and new trending #CO2 {social media applications}, as well as #IL2 {digital/online journalism will also be covered}. Weekly readings will be assigned. #CO3 {CMC and group communication}. Students will write and present a final paper demonstrating their learning and knowledge of the course related materials</p>	Information and data literacy {#IL1, #IL2} Communication and collaboration {#CO1, #CO2, #CO3}
Interaction design and virtual reality/(H)	<p>This course will #IL1 {cover the process (e.g., design-build-iterate loop)} and #CA1 {skills to build innovative interactive systems}, #IL2 {with a light touch on usability testing}. The process section covers idealization, #CO1 {lo-fi prototyping and heuristic evaluation}. The skill section includes practical knowledge, techniques and #CA2 {tools in creating interactive systems} in particularly the #CO2 {virtual reality applications}. In addition, this course teaches basic #CO3 {understanding to interaction design by walking students through common practices/principles in the design of multitouch, tangible, and body based user interfaces}. Students will form #CO4 {project teams} to #CA3 {create novel interactive devices and showcase the derived interactions}</p>	Information and data literacy: {#IL1, #IL2} Communication: {#CO1, #CO2, #CO3, #CO4} Content creation: {#CA1, #CA2, #CA3}

The classification model was trained with each ML algorithm using nine of the ten subsets (the training folds), with the remaining subsets being used for validation (verification folds). The average results of ten iterations for each ML classifier were included in the analysis to prevent overfitting.

Model evaluation

To identify the performances of the ML models, accuracy, precision, sensitivity, F1-score, and kappa were considered. Accuracy was measured by the percentage of correctly classified syllabi. Precision reflected the ability of the ML model to identify only the relevant syllabus for each DCPL. Sensitivity measured where a certain syllabus for a DCPL should have been classified and how many times it was correctly classified. The F1-score was established by the weighted average of precision and sensitivity and provided an overall metric for evaluating the ML classifier. All of the above metrics give values between 0 and 1. Values that approach 0 indicate increasingly unacceptable performance, while values approaching 1 indicate increasingly excellent performance. Kappa was used to evaluate the consistency between the ML results and the human classification results in a range from -1 to 1. Higher values indicate better agreement between the assessments.

Results

The classification effectiveness of the ML models

In total, 548, 303, 139, and 210 courses were labeled DCPL=H, DCPL=M, DCPL=L, and DCPL=N, respectively. We used the test set to evaluate the performances of the ML models and found that the accuracy levels of the four classification models, i.e., SVM, KNN, logit, and NB, ranged from 0.768 to 0.922, while the kappa value ranged from 0.662 to 0.886. Thus, TD-IDF demonstrated high accuracy and consistency as a feature extraction method for syllabus classification. We accordingly evaluated the ML models using tenfold cross-validation. The tenfold cross-validation results (Table 3) showed that the accuracy for the four classifiers ranged from 0.656 to 0.713, and precision ranged from 0.594 to 0.712, with sensitivities ranging from 0.587 to 0.708. These results indicated that more than 71% of the syllabi were correctly classified according to their DCPL using ML models. This was better than expected, as university syllabi usually do not have fixed norms. A possible reason for the success may be that the TD-IDF identified the difference and similarity in documents primarily through word frequency, and more representative distinguishable keywords simplified the distinctions among documents (McHugh et al. 2020). The syllabus describes the use of digital competence skills and tools, often including unique terms such as internet, e-mail, upload, software simulation, and source code (Stanny et al. 2015). These specific words can be distinguished from general words so that their significant features can be extracted to enhance the classification ability.

Precision and sensitivity should also be considered when educational stakeholders review the digital competence in university syllabi. Many universities are providing digital-competence development programs for faculty to prepare them for dealing with possible learning environment transitions that may occur at any time. As a result, institutions often investigate whether the digital competence levels of their faculty have increased or changed after a certain period following a training session. A high-precision ML model can provide an efficient and effective means for the institution to address the issue above. More specifically, the greater the ML model precision, the better all courses can be classified into corresponding categories. Here, institutions can generate comprehensive reports on the ratio of courses at each level of digital competence with the changes that have been made. Another potential action would be to seek universal digital competence development on the campus by prioritizing the assignment of training resources to faculty, students, and departments that currently demonstrate lower levels of digital competence. Here, a highly sensitive classifier will be particularly appropriate for addressing this issue. This is because the more sensitive the model is, the more that courses that do not significantly integrate digital competence can be correctly distinguished from those that have not integrated digital competence at all. In this way, resources to support departments, faculty, and students as they develop higher digital competence levels can be precisely delivered.

Agreement between ML models and human classification

Beyond the accuracy, precision, and sensitivity of the ML models, we also evaluated the agreement between human and ML models using kappa as a measure of agreement between behavioral observers. Kappa is often used in ML models to compare agreement between machine and human judgment. One study showed that the average kappa for SVM, KNN, logit, and NB, using tenfold cross-validation, ranged from 0.460 to 0.555. Values of kappa higher than 0.4 are desirable when comparing human and machine evaluations (Sakiyama et al. 2008). That is, the results of using ML models for syllabus classification are consistent with those for human classification. Our results showed that, based on the confusion matrix (Fig. 1) for the four classifiers, the syllabi for each DCPL were mostly correctly differentiated. We then randomly selected syllabi identified as having no, low, moderate, and high DCPLs

Table 3 Evaluation of classification models

	Test set validation					Tenfold cross-validation				
	ACC	Pre	Sen	F1	Kappa	ACC	Pre	Sen	F1	Kappa
SVM	0.920	0.920	0.921	0.920	0.882	0.713	0.712	0.708	0.648	0.555
Logit	0.921	0.923	0.922	0.922	0.885	0.657	0.638	0.638	0.464	0.460
KNN	0.922	0.930	0.922	0.922	0.886	0.656	0.629	0.593	0.598	0.451
NB	0.768	0.772	0.768	0.767	0.664	0.656	0.594	0.587	0.589	0.523

and examined their contents. The syllabi classified by the ML model included descriptions related to the assessment criteria involved in this study. For example, the syllabus for *Engineering Mathematics* (DCPL=N) had no words related to digital skills or tools. Students who took the *Anthropocene* (DCPL=L) course were required to access information on a website, referring to the area of information and data literacy. Similarly, sharing content via digital media and communicating and collaborating was required for students to take the Bilingual Creative Writing (DCPL=M) course. Likewise, students who took the Computer Programming and Engineering Application (DCPL=H) course were expected to be taught to create their own content using a programming language. These relate to digital content creation. This result provides further evidence that ML can be used to classify syllabi in manner that agrees well with human evaluation results (Table 4).

Discussion

Using ML text categorization to analyze syllabus

This study was conducted in response to two research questions, as follows. RQ1: How well can ML evaluate the level of digital competence applied in a course from its syllabus? RQ2: How well can ML classify the levels of digital competence present in courses relative to human evaluation? We marked 1200 syllabus to train the ML models to perform classification to address these questions. The results show that the four ML models used demonstrated high performance on the classification task. ML methods can be used to evaluate the level of digital competence integrated into a course with high accuracy, precision, and sensitivity. For the second research question, our results showed that the ML models could identify syllabi that covered different levels of digital competence and could produce classifications that were highly consistent with those produced by human beings.

The results of this study echoed previous findings that there is a large amount of qualitative data in the field of education and that applying ML methods to text classification can provide accurate, consistent, relevant, and verifiable results to facilitate educational data analysis (Immonen et al. 2015). This study also suggested that even though university syllabi describe content in different domains with unstructured text, ML methods can still provide a reliable, effective, and efficient means of automatically evaluating the level of digital competence integrated into a course. Of the four most popular ML models used for text classification, our results indicate that SVM has the best performance on the classification task and achieves the highest accuracy (0.713) and agreement (0.555) in tenfold cross-validation. This result corresponds to previous findings that indicated that SVM may perform better than other classifiers in text classification tasks that do not feature datasets that are very large (< 6000 instances) (Yu and Xu 2008). Therefore, we suggest that SVM be taken as a benchmark for developing further algorithms to reinforce the performance and consistency of the classification model in future studies. It is worth noting that high and moderate DCPL courses were misclassified more often than other levels.

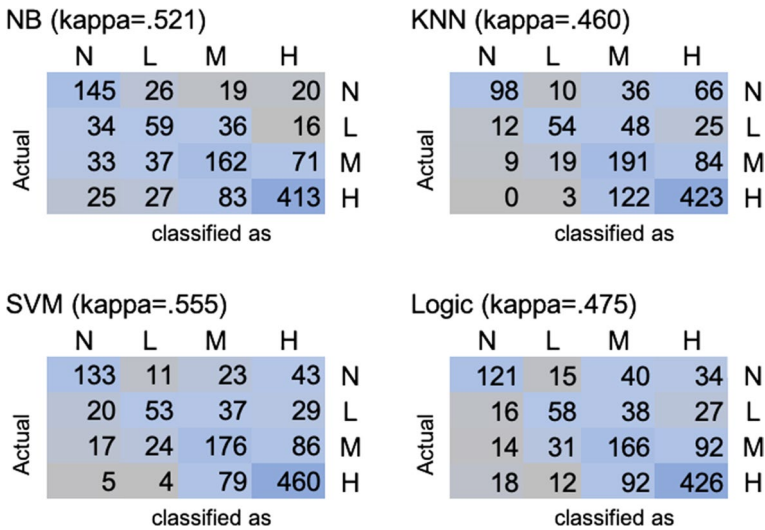


Fig. 1 The average confusion matrix obtained through tenfold cross validation

This may be due to an overlap in the definitions and assessment rules for moderate and high digital competence courses relative to the low and no DCPL courses. For example, interaction through technologies is represented by communication and collaboration, which indicates a moderate level, and content development is related to the domain of digital content creation and entails a high DCPL level. However, both of these levels call for students to be able to use various digital tools to accomplish learning tasks, and many digital tools and terms used in syllabi overlap or are similar, such as PowerPoint, electronics performance, the use of computers, or reliance on learning systems. Because the text document classification relied on extracting significant features from keywords, these similarities may have contributed to misclassification instances by ML models when evaluating syllabi. Further research is needed to address this issue.

Implications for evaluating digital competence based on ML models for higher education

The issue of digital competence in higher education institutions has attracted particular attention in recent years (König et al. 2020). For the purpose of professional knowledge training and career preparation, students’ learning and use of digital technologies during higher education are crucial for their learning and life (Tsankov and Damyanov 2017). Understanding how and to what extent university instructors integrate digital competencies into their classes has become an essential issue. However, past methods of evaluating competence have critical limitations that need to be improved (Zhao et al. 2021). The results of this study confirm that the use of machine learning assessment methods is not only feasible but also highly consistent

Table 4 Example of result of syllabus classified by ML models

Course name/(DCPL)	Part of the syllabus	Descriptions rated to DCPL
Engineering mathematics/(N)	<p>Linear Algebra. Matrices, Vectors, Determinants, Linear systems Linear Algebra. Matrix Eigenvalue Problems Vector Differential Calculus. Grad, Div, Curl Vector Integral Calculus. Integral Theorems Fourier series, Integrals, and Transforms</p>	N/A
Anthropocene/(L)	<p>This course explores a concept: the Anthropocene, which has been proposed as a new planetary epoch in which human processes drive all major earth systems. Under the banner of the Anthropocene, actors from a great variety of intellectual contexts are beginning to find a new language to talk about planetary impacts. All changes will be announced and explained in class and updated on the #IP1 {course web site: https://googl/59GvaE}</p>	Information and data literacy: #PI
Bilingual creative writing/(M)	<p>This course introduces writing techniques that aim to transcend genre by learning some tricks and common practices in creative writing cultures. Students from music and media arts backgrounds are encouraged to present final projects and (at least some) assignments in non-textual media such as #CO1 {sound files, short films, mock presentations (for instance, using the format of a PowerPoint, Ted Talk, or podcast)}, or #IL {even (live computer music) or acoustic music with sampled language #CO2 {in electronics performances}}, to name a few possibilities (designed to encourage interdisciplinary synergy)</p>	<p>Information and data literacy: #PI Communication and collaboration: {#CO1, #CO2}</p>
Computer programming and engineering application/(H)	<p>This course is intended for both non-computer science major students who do not have extensive experience in computer programming and system design engineers who are interested in integrating code. From the stage of system analysis and design, the #CA1 {students will begin to learn the concept of object-oriented programming (OOP)} and #CA2 {solve the engineering problem using C/C++}. By using the material of each topic, the #CA3 {student will cultivate good programming concept and methodology}. Finally, there will be a project in which #CO1 {students are required to use version control system and team work to complete the assignment}. In this case, #CA4 {students can sharpen their programming skills from problem solving, and also lay the groundwork for advanced programming language}...</p>	<p>Communication and collaboration: {#CO1} Digital content creation: {#CA1, #CA2, #CA3, #CA4}</p>

with human assessment results, which has several implications for assessing digital competencies in higher education institutions. First, the benefits of automated processing by machines provide a solution to the current human and time constraints that universities face in investigating digital competencies on campus. This high-performance approach can easily be applied to answer questions related to digital competence in higher education institutions in real time. For example, are universities offering sufficient courses to develop digital competencies? What percentage of students are enrolled in courses that cover digital competencies? In addition, from a pedagogical practice perspective, the syllabus is one of the documents that demonstrate teachers' technological pedagogical content knowledge (TPACK), and the solution proposed in this study can be used to provide objective evidence of teacher development theory and practice (Loveless 2011). For example, how do university instructors respond to the current digital age in terms of their technological knowledge? What levels of digital competence do instructors demonstrate in courses and in what ways? In other words, higher educational institutes can use this approach to gain a comprehensive understanding of the digital competency profile of campus faculty and students; prepare to provide appropriate training programs or assistance to students, faculty, or departments; and explore strategies to enhance faculty members' and students' digital competency levels based on evidence. Figure 2 provides the conceptual framework.

Conclusion and limitations

The ways in which higher education instructors integrate digital competence into their pedagogy are essential to obtaining a quality education and are a significant pathway for students to gain digital skills, especially during the present COVID-19 crisis. Hence, research to understand digital competence in higher education from different perspectives is urgently needed. Previous studies have focused on applying self-reported, time-consuming, or human-intensive methods to assess digital competence. This study adopted ML methods to evaluate syllabi and propose solutions for assessing the degree of integration of digital competence in university courses. The results of this study suggest that the solution proposed in this study is efficient, effective, and objective relative to conventional methods of assessing digital competence. In addition, higher education institutes can more efficiently assess and practice digital competence and develop educational interventions. Taking this practical approach, universities can direct resources to increase their digital competence levels more effectively and plan for current and future development.

Although the results of this study are promising, they should be interpreted in light of its limitations. First, the methodology of this study produced the desired results; however, this study focused on the TF-IDF algorithm, and contextual sentences, semantic considerations, and implicit meaning were not considered. It is worth considering the long short-term memory model for advanced natural language processing to reinforce analyzing performance and minimize the gap between machines and humans evaluating digital competence in the future. Second, this study was conducted at a research university, and whether the effectiveness of this

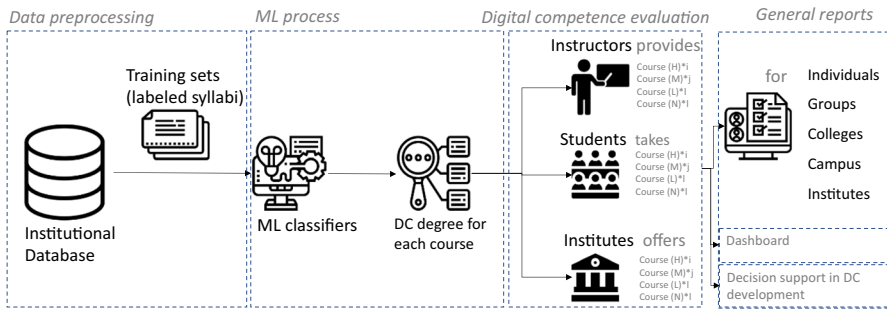


Fig. 2 The conceptual framework of analysis the digital competence in university

method for other types of universities (e.g., teaching universities and comprehensive universities) is consistent with the results at research universities is unclear. Future research could attempt to confirm the generalizability of this method by demonstrating the differences in its effectiveness across university types. Finally, there might still be a digital divide between instructors and students, and it would be valuable if future research further uncovered the differences between teachers' and students' digital competency levels to improve this issue.

Acknowledgements This research was supported by the Higher Education Sprout Project of National Yang Ming Chiao Tung University (NYCU) and the Ministry of Education (MOE), Taiwan, as well as the Ministry of Science and Technology in Taiwan through Grant Numbers MOST 108-2511-H-009-019-MY2 and 111-2410-H-A49-029. We would also thank Yu Chen Wu for her support during the study.

Author contribution The single authors of the current article contributed to the conception or design of the work, data analysis, and manuscript preparation and submission.

Funding This research was supported by the Higher Education Sprout Project of National Yang Ming Chiao Tung University (NYCU) and the Ministry of Education (MOE), Taiwan, as well as the Ministry of Science and Technology in Taiwan through Grant Numbers MOST 108-2511-H-009-019-MY2 and 111-2410-H-A49-029.

Data availability The data used in this study can be found on the NYCU website (<https://timetable.nycu.edu.tw/>).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study was approved by the Institutional Review Board B (IRB-B) of National Yang Ming Chiao Tung University.

Informed consent This article does not contain any studies with human participants performed by any author.

References

- Alarcón R, del Jiménez E, de Vicente-Yagüe MI (2020) Development and validation of the DIGIGLO, a tool for assessing the digital competence of educators. *Br J Educ Technol* 51(6):2407–2421. <https://doi.org/10.1111/bjct.12919>
- Al-Rimy BAS, Maarof MA, Alazab M, Alsolami F, Shaid SZM, Ghaleb FA, Al-Hadhrami T, Ali AM (2020) A pseudo feedback-based annotated TF-IDF technique for dynamic crypto-ransomware pre-encryption boundary delineation and features extraction. *IEEE Access* 8:140586–140598. <https://doi.org/10.1109/ACCESS.2020.3012674>
- Alsmadi I, Hoon GK (2019) Term weighting scheme for short-text classification: Twitter corpuses. *Neural Comput Appl* 31(8):3819–3831. <https://doi.org/10.1007/s00521-017-3298-8>
- Antón-Sancho Á, Vergara D, Fernández-Arias P (2021) Self-assessment of soft skills of university teachers from countries with a low level of digital competence. *Electronics* 10(20):2532
- Bensen H, Silman F (2012) Three preparatory schools' syllabus designs in the Turkish republic of North-Cyprus. *Eurasian J Educ Res*. <https://eric.ed.gov/?id=EJ1057357>
- Bartolomé J, Garaizar P, Larucea X (2022) A pragmatic approach for evaluating and accrediting digital competence of digital profiles: a case study of entrepreneurs and remote workers. *Technol Knowl Learn* 27(3):843–878
- Barthakur A, Joksimovic S, Kovanovic V, Mello RF, Taylor M, Richey M, Pardo A (2022) Understanding depth of reflective writing in workplace learning assessments using machine learning classification. *IEEE Trans Learn Technol*. <https://doi.org/10.1109/TLT.2022.3162546>
- Beuoy M, Boss K (2019) Revealing instruction opportunities: a framework-based rubric for syllabus analysis. *Ref Serv Rev* 47(2):151–168. <https://doi.org/10.1108/RSR-11-2018-0072>
- Blau I, Shamir-Inbal T, Avdiel O (2020) How does the pedagogical design of a technology-enhanced collaborative academic course promote digital literacies, self-regulation, and perceived learning of students? *Internet Higher Educ* 45:100722. <https://doi.org/10.1016/j.iheduc.2019.100722>
- Boss K, Drabinski E (2014) Evidence-based instruction integration: a syllabus analysis project. *Ref Serv Rev* 42(2):263–276. <https://doi.org/10.1108/RSR-07-2013-0038>
- Cabero-Almenara J, Barroso-Osuna J, Gutiérrez-Castillo J-J, Palacios-Rodríguez A (2021) The teaching digital competence of health sciences teachers: a study at Andalusian universities (Spain). *Int J Environ Res Public Health* 18(5):2552. <https://doi.org/10.3390/ijerph18052552>
- Caena F, Redecker C (2019) Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (Digcompedu). *Eur J Educ* 54(3):356–369. <https://doi.org/10.1111/ejed.12345>
- Carretero S, Vuorikari R, Punie Y (2017) The digital competence framework for citizens. Publications Office of the European Union.
- Chiu T-F, Chu D, Huang S-J, Chang M, Liu Y, Lee JJ (2021) Facing the coronavirus pandemic: an integrated continuing education program in Taiwan. *Int J Environ Res Public Health* 18(5):2417. <https://doi.org/10.3390/ijerph18052417>
- Domínguez-Lloria S, Fernández-Aguayo S, Marín-Marín J-A, Alvarinas-Villaverde M (2021) Effectiveness of a collaborative platform for the mastery of competencies in the distance learning modality during COVID-19. *Sustainability* 13(11):5854. <https://doi.org/10.3390/su13115854>
- Dubicki E (2019) Mapping curriculum learning outcomes to ACRL's Framework threshold concepts: a syllabus study. *J Acad Librariansh* 45(3):288–298. <https://doi.org/10.1016/j.acalib.2019.04.003>
- Ferrari A (2013) DIGCOMP: A framework for developing and understanding digital competence in Europe. Publications Office of the European Union Luxembourg.
- Golowko N (2021) The improvement of sustainable employability transfer in higher education institutions using large scale data bases and machine learning. In: Golowko N (ed) *Future skills in education: knowledge management AI and sustainability as key factors in competence-oriented education*. Springer, pp 165–185. https://doi.org/10.1007/978-3-658-33997-5_6
- Gower R, Walters S, Phillips D (1983) *Teaching practice handbook*. Heinemann, London
- Guillén-Gámez FD, Mayorga-Fernández MJ (2020) Quantitative-comparative research on digital competence in students, graduates and professors of faculty education: an analysis with ANOVA. *Educ Inf Technol* 25(5):4157–4174. <https://doi.org/10.1007/s10639-020-10160-0>
- Heidari E, Mehrvarz M, Marzooghi R, Stoyanov S (2021) The role of digital informal learning in the relationship between students' digital competence and academic engagement during the COVID-19 pandemic. *J Comput Assist Learn* 37(4):1154–1166. <https://doi.org/10.1111/jcal.12553>

- Hernández-Martín A, Martín-del-Pozo M, Iglesias-Rodríguez A (2021) Pre-adolescents' digital competences in the area of safety: does frequency of social media use mean safer and more knowledgeable digital usage? *Educ Inf Technol* 26(1):1043–1067. <https://doi.org/10.1007/s10639-020-10302-4>
- Iatrellis O, Savvas IK, Kameas A, Fitsilis P (2020) Integrated learning pathways in higher education: a framework enhanced with machine learning and semantics. *Educ Inf Technol* 25(4):3109–3129. <https://doi.org/10.1007/s10639-020-10105-7>
- Immonen A, Pääkkönen P, Ovaska E (2015) Evaluating the quality of social media data in big data architecture. *IEEE Access* 3:2028–2043. <https://doi.org/10.1109/ACCESS.2015.2490723>
- Joachims T (1998) Text categorization with Support Vector Machines: Learning with many relevant features. In: Nédellec C, Rouveirol C (eds) *Machine Learning: ECML-98*. Springer, New York, pp 137–142. <https://doi.org/10.1007/BFb0026683>
- Kadhimi AI (2019) Survey on supervised machine learning techniques for automatic text classification. *Artif Intell Rev* 52(1):273–292. <https://doi.org/10.1007/s10462-018-09677-1>
- Kawasaki H, Yamasaki S, Masuoka Y, Iwasa M, Fukita S, Matsuyama R (2021) Remote teaching due to COVID-19: an exploration of its effectiveness and issues. *Int J Environ Res Public Health* 18(5):2672. <https://doi.org/10.3390/ijerph18052672>
- König J, Jäger-Biela DJ, Gluttsch N (2020) Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *Eur J Teach Educ* 43(4):608–622. <https://doi.org/10.1080/02619768.2020.1809650>
- Lucas M, Bem-Haja P, Siddiq F, Moreira A, Redecker C (2021) The relation between in-service teachers' digital competence and personal and contextual factors: What matters most? *Comput Educ* 160:104052. <https://doi.org/10.1016/j.compedu.2020.104052>
- Esteve-Mon F, Adell-Segura J, Ángeles Llopis Nebot M, Valdeolivas Novella G, Pacheco Aparicio J (2019) The development of computational thinking in student teachers through an intervention with educational robotics. *J. Inf. Technol. Educ. Innov. Pract.* 18:139–152. <https://doi.org/10.28945/4442>
- Madsen SS, Thorvaldsen S, Archard S (2018) Teacher educators' perceptions of working with digital technologies. *Nordic J Digit Lit* 13(03):177–196. <https://doi.org/10.18261/issn.1891-943x-2018-03-04>
- Manca S, Delfino M (2021) Adapting educational practices in emergency remote education: Continuity and change from a student perspective. *Br J Edu Technol* 52(4):1394–1413. <https://doi.org/10.1111/bjet.13098>
- Mattar J, Ramos DK, Lucas MR (2022) digcomp-based digital competence assessment tools: literature review and instrument analysis. *Educ Inf Technol* 1:1–25
- McHugh D, Shaw S, Moore TR, Ye LZ, Romero-Masters P, Halverson R (2020) Uncovering themes in personalized learning: using natural language processing to analyze school interviews. *J Res Technol Educ* 52(3):391–402. <https://doi.org/10.1080/15391523.2020.1752337>
- Monroy García FA, Llamas-Salguero F, Fernández-Sánchez MR, Carrión del Campo JL (2020) Digital technologies at the pre-university and university levels. *Sustainability* 12(24):10426. <https://doi.org/10.3390/su122410426>
- Mowafy M, Rezk A, El-Bakry H (2018) An efficient classification model for unstructured text document. *Am J Comput Sci Inf Technol.* <https://doi.org/10.21767/2349-3917.100016>
- Olszewski B, Crompton H (2020) Educational technology conditions to support the development of digital age skills. *Comput Educ.* <https://doi.org/10.1016/j.compedu.2020.103849>
- Parkes J, Harris MB (2002) The Purposes of a syllabus. *Coll Teach* 50(2):55–61. <https://doi.org/10.1080/87567550209595875>
- Pinto M, Fernández-Pascual R, Caballero-Mariscal D, Sales D (2020) Information literacy trends in higher education (2006–2019): visualizing the emerging field of mobile information literacy. *Scientometrics* 124(2):1479–1510. <https://doi.org/10.1007/s11192-020-03523-4>
- Richardson V (1990) Significant and worthwhile change in teaching practice. *Educ Res* 19(7):10–18
- Saide S, Sheng ML (2021) Knowledge exploration–exploitation and information technology: crisis management of teaching–learning scenario in the COVID-19 outbreak. *Technol Anal Strategic Manag* 33(8):927–942. <https://doi.org/10.1080/09537325.2020.1854714>
- Sakiyama Y, Yuki H, Moriya T, Hattori K, Suzuki M, Shimada K, Honma T (2008) Predicting human liver microsomal stability with machine learning techniques. *J Mol Graph Model* 26(6):907–915. <https://doi.org/10.1016/j.jmglm.2007.06.005>

- Sánchez-Cruzado C, Santiago Campión R, Sánchez-Compañía MT (2021) Teacher digital literacy: the indisputable challenge after COVID-19. *Sustainability* 13(4):1858. <https://doi.org/10.3390/su13041858>
- Schina D, Esteve-González V, Usart M, Lázaro-Cantabrana J-L, Gisbert M (2020) The integration of sustainable development goals in educational robotics: a teacher education experience. *Sustainability* 12(23):10085. <https://doi.org/10.3390/su122310085>
- Sillat LH, Tammets K, Laanpere M (2021) Digital competence assessment methods in higher education: a systematic literature review. *Educ Sci* 11(8):402
- Stanny C, Gonzalez M, McGowan B (2015) Assessing the culture of teaching and learning through a syllabus review. *Assess Eval High Educ* 40(7):898–913. <https://doi.org/10.1080/02602938.2014.956684>
- Starkey L (2020) A review of research exploring teacher preparation for the digital age. *Camb J Educ* 50(1):37–56. <https://doi.org/10.1080/0305764X.2019.1625867>
- Tsankov N, Damyanov I (2017) Education majors' preferences on the functionalities of e-learning platforms in the context of blended learning. *Int J Emerg Technol Learn (IJET)* 12:202. <https://doi.org/10.3991/ijet.v12i05.6971>
- Väätäjä JO, Ruokamo H (2021) Conceptualizing dimensions and a model for digital pedagogy. *J Pac Rim Psychol* 15:1834490921995395. <https://doi.org/10.1177/1834490921995395>
- Vorobel O, Voorhees TT, Gokcora D (2021) Language learners' digital literacies: focus on students' information literacy and reading practices online. *J Comput Assist Learn* 37(4):1127–1140. <https://doi.org/10.1111/jcal.12550>
- Weber H, Hillmert S, Rott KJ (2018) Can digital information literacy among undergraduates be improved? Evidence from an experimental study. *Teach High Educ* 23(8):909–926. <https://doi.org/10.1080/13562517.2018.1449740>
- Xu S (2018) Bayesian Naïve Bayes classifiers to text classification. *J Inf Sci* 44(1):48–59. <https://doi.org/10.1177/0165551516677946>
- Yakubu MN, Abubakar AM (2021) Applying machine learning approach to predict students' performance in higher educational institutions. *Kybernetes* 51(2):916–934
- Yu B, Xu Z (2008) A comparative study for content-based dynamic spam classification using four machine learning algorithms. *Knowl Based Syst* 21(4):355–362. <https://doi.org/10.1016/j.knosys.2008.01.001>
- Zhao Y, Pinto Llorente AM, Sánchez Gómez MC (2021) Digital competence in higher education research: A systematic literature review. *Comput Educ* 168:104212. <https://doi.org/10.1016/j.compedu.2021.104212>
- Zhou J (2021) The role of libraries in distance learning during COVID-19. *Inf Dev*. <https://doi.org/10.1177/02666669211001502>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.