Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers

Cheng-Huan Chen
Asia University, Taiwan

Stephen J. H. Yang, Jian-Xuan Weng
National Central University, Taiwan

Hiroaki Ogata
Kyoto University, Japan

Chien-Yuan Su
National University of Tainan, Taiwan

Providing early predictions of academic performance is necessary for identifying at-risk students and subsequently providing them with timely intervention for critical factors affecting their academic performance. Although e-book systems are often used to provide students with teaching/learning materials in university courses, seldom has research made the early prediction based on their online reading behaviours by implementing machine learning classifiers. This study explored to what extent university students' academic achievement can be predicted, based on their reading behaviours in an e-book supported course, using the classifiers. It further investigated which of the features extracted from the reading logs influence the predictions. The participants were 100 first-year undergraduates enrolled in a compulsory course at a university in Taiwan. The results suggest that logistic regression Gaussian naïve Bayes, supports vector classification, decision trees, and random forests, and neural networks achieved moderate prediction performance with accuracy, precision, and recall metrics. Furthermore, the Bayes classifier identified almost all at-risk students. Additionally, student online reading behaviours affecting the prediction models included: turning pages, going back to previous pages and jumping to other pages, adding/deleting markers, and editing/removing memos. These behaviours were significantly positively correlated to academic achievement and should be encouraged during courses supported by e-books.

Implications for practice or policy:
- For identifying at-risk students, educators could prioritise using Gaussian naïve Bayes in an e-book supported course, as it shows almost perfect recall performance.
- Assessors could give priority to logistic regression and neural networks in this context because they have stable achievement prediction performance with different evaluation metrics.
- The prediction models are strongly affected by student online reading behaviours, in particular by locating/returning to relevant pages and modifying markers.

Keywords: machine learning classifier, machine learning classification algorithm, academic achievement, reading behaviour, e-book system, early prediction, at-risk student

Introduction

In recent years considerable interest has emerged in the domain of education research regarding the application of artificial intelligence (Chen et al., 2020; Hwang et al., 2020; Yang et al., 2021). One development addresses how predicting students' performance has become an essential research topic in different academic disciplines (e.g., Lu et al., 2018; Tsai et al., 2020), and the goal of being able to predict levels for students learning performance, score, ability or mark from specific information, aspects or behaviour of those students (Romero & Ventura, 2013). For example, how instructors track behaviour patterns and find student diversity to enable them to apply appropriate educational approaches (Xenos, 2004), or how instructors should try to identify high-risk students in a cohort, identify possible dropouts,

or failed course students, and prevent this from happening (Heuer & Breiter, 2018; Wakelam et al., 2019; Wu et al., 2020). Instructors may use a range of remedial or corrective strategies to interact with the at-risk students and provide them assistance to enhance their course learning performance.

It is expected that applying machine learning algorithms to predict student academic performance in e-learning research is likely to become more common. Recently, several classic machine learning algorithms have been used to build classification models for student performance prediction, including logistic regression (LR), Gaussian naïve Bayes (GaussianNB), support vector classification (SVC), decision tree (DT), random forest (RF), and neural network (NN) (Huang et al., 2020; Luan & Tsai, 2021; Wu et al., 2020). LR and GaussianNB are linear classifiers and statistical classification methods. SVC is vector space-based classification and a kernel-based classifier that can be linear or nonlinear (Murty & Raghava, 2016). DT and RF are tree-based classification methods. The advantage of these technologies is in data-driven and evidence-based research as opposed to model-driven. That is, as in model-based linear or non-linear methods, the stated methods do not assume that there is an explicit relational model between the data. Instead, the model structure and model parameters they use are derived from the actual data set of the problem (Kitto et al., 2017; Lykourentzou et al., 2009). Currently, a variety of predictive modeling techniques (Tomasevic et al., 2020) are used to predict students' performance in a course and identify those that are at-risk (Jin et al., 2011; Wu et al., 2020).

Prediction model using machine learning techniques

The machine learning techniques using algorithms that derive general assumptions from externally provided examples will make predictions for future examples (Kotsiantis et al., 2010). Nowadays, machine learning techniques can be used to predict students' academic performance based on predicted variable characteristics (Hämäläinen & Vinni, 2010), such as classification (the predictor is a categorical value), regression (the predictor is a continuous value), or density estimation (the predictor is a probability density function). This occurs by examining student e-portfolio records or their operational progress, retrieved from system logs. This process has made significant progress in recent years (Ashraf et al., 2018; Cheng, 2017; Lykourentzou et al., 2009) in capturing rich data trails (Siemens & Long, 2011; Wen et al., 2018). For example, some researchers have adopted students' learning experience and previous computer training as a pre-entry variable to predict which learners have completed university-level online distance education courses (Dupin-Bryant, 2004; Xenos et al., 2002). Morris et al. (2005) developed a classification algorithm to forecast college students' withdrawal from or retention in an online course based on their GPA and SAT mathematics score. They reported an accurate classification rate of 74.5%. Xenos (2004) utilised Bayesian networks to model past educational experiences of 800 university students' behaviours by analysing both past student data and data related to the students' progress through the academic year. Their model was presented to help teachers determine the reasons or indicators for student dropouts, and further help them make decisions regarding their educational procedure. Similarly, Bekele and McPherson (2011) created a Bayesian prediction model to determine the strengths of personality traits in predicting the mathematical performance of Addis Ababa high school students.

Moreover, some research used different or multimodal machine learning techniques (i.e., integration of two or more machine learning algorithms) to interpret various forms of data to automatically evaluate complex learning processes (Di Mitri et al., 2017), or to automatically predict the learning performance based on multimodal data (Giannakos et al., 2019) to cope with complex and changing learning contexts to support education administrators and tutors in making decisions under conditions of uncertainty (Xenos, 2004). For example, Kotsiantis et al.'s (2003) study, compared a variety of machine learning techniques (i.e., DT, NN, LR, naïve Bayes [NB], SVM, and instance-based learning algorithms) to predict students' trends to dropping a university class. Their study used a plenary attendance and written assignments dataset regarding the performance of 354 students. Huang and Fang (2013) developed a set of 24 validated mathematical models integrated multiple linear regression, multilayer perception network, radial basis function network, and support vector machine, and six combinations of predictor variables based on the dataset collected from 323 undergraduates in four semesters in order to predict students' grades for an engineering dynamics course. Their results demonstrated the value of using performance data gathered during the semester for predictive purposes. Mayilvaganan and Kalpanadevi (2014) employed classification techniques to predict the range of students' academic performance, such as predicting very good learners, good learners, middle learners, and slow learners from academic datasets. Marbouti et al. (2015) implemented three LR-based models to predict at-risk students based on their homework, quiz, and

attendance records in a large first-year engineering course, 3 times during the semester. The models were optimised for identifying at-risk students, and were able to identify at-risk and successful students with overall accuracy of 79%, 90%, and 98%, respectively. Marbouti et al. (2016) proposed predictive methods that integrated an aggregate model consisting of support vector machine, *k*-nearest neighbours, and NB classifier to forecast at-risk students in a course that used standards-based grading. Using this method was they were able to identify high-risk students early, and inform both the instructors and the students. Tomasevic et al. (2020) compared supervised machine learning techniques aimed to predict students' exam performance, discover students at a high risk of dropping out of the course, and predict their future achievements. They concluded that adopting artificial NNs by processing the student engagement data and past performance data demonstrated the highest precision, but the usage of demographic data did not present a significant influence on the precision of predictions.

Student reading behaviours in an e-book system

Thanks to technological progress, many electronic learning platforms or systems can not only provide a variety of learning materials or course materials for learners to learn autonomously, but can also automatically collect student learning data, including student achievements and patterns of behaviour. For example, an e-book system is often used to provide students with teaching or learning materials in a university course, enabling them to carry out pre- or post-class reading (Chen et al., 2020). It stores digital teaching and learning materials. It also records students' reading behaviours toward studying digital materials.

An e-book system, which supports bookmarks, highlights, annotations, and queries, built by Kyoto University, was used in this study. Students use these e-book features for their reading and learning (Akçapınar et al., 2019; Ogata et al., 2015). Ideally, students' reading logs and highlights (Yang et al., 2020) can be recorded on the system server, allowing better investigation and understanding of the relationship between student reading behaviours toward studying digital materials used in lectures and their academic achievement. Empirical and practical research is thus encouraged to deal with this issue in the context of college students from online reading behaviour perspectives. In addition, the correlation between students' online reading behaviours in an e-book system and their academic achievement should also be investigated so as to identify the key e-book features that may affect students' performance (Chen & Su, 2019) and engagement (Flanagan et al., 2020).

Purpose and research questions

Although several previous studies have employed machine learning techniques to predict students' academic performance, few studies involving students' viewing and operation records in an e-book system have been done. This study therefore used machine learning classification algorithms, to better understand the extent to which they can predict university students' academic performance in the course, based on their reading logs recorded in an e-book system. These logs were used to identify at-risk students at an earlier stage. In this study, machine learning classifiers including logistic regression (LR), Gaussian naïve Bayes (GaussianNB), support vector classification (SVC), decision tree (DT), random forest (RF), and neural network (NN) were used. Additionally, key factors that affect the accuracy of academic performance prediction were also explored. Accordingly, this study aimed to answer the following research questions and hypotheses:

RQ1. In a university course supported by an e-book system, to what extent can the academic achievement of students be predicted based on their reading logs, by using machine learning classification algorithms?

    H1.1: The LR classifier performs well based on the online reading logs of university students.
    H1.2: The GaussianNB classifier performs well based on the online reading logs of university students.
    H1.3: The SVC classifier performs well based on the online reading logs of university students.
    H1.4: The DT classifier performs well based on the online reading logs of university students.
    H1.5: The RF classifier performs well based on the online reading logs of university students.
    H1.6: The NN classifier performs well based on the online reading logs of university students.

RQ2. Corresponding to the previous question, what reading behaviours can affect the predictions of student academic achievement?

H2: University students' online reading behaviours in the e-book system correlate with their academic achievement in the prediction model.

Method

Participants and settings

In order to ensure the robustness of prediction results, the researchers obtained two datasets from two different classes for this study. The findings of one would be considered robust if confirmed in the other. Two classes at a university in Taiwan, containing a total of 100 first-year undergraduates between the ages of 18 and 19 were invited to participate in this study. One class consisted of 53 students: 12 were male and 41 were female. The other class was composed of 47 students: 24 were male and 23 were female. The average scores for the two classes on an information proficiency test conducted by the university prior to the students' enrolment in the classes did not differ significantly, $t(90) = -1.38$, $p = .17$.

All participants were enrolled in the second half of the same 2-credit mandatory course, Information and Technology, covering seven units: (1) free web hosting services, (2) introduction to computer basics, (3) word processing software, (4) presentation software, (5) spreadsheet, (6) cloud infrastructure and services, (7) operating system and task manager. The slides for each unit were uploaded to BookRoll (Version 2.3), which is a non-commercial, digital teaching material delivery system and also an e-book system that allows students to view digital materials used in lectures (Ogata et al., 2017; Ogata et al., 2015). Both classes had the same content and used BookRoll to provide the same materials for the lecture slides. The researchers obtained consent from the participating university and the class teacher (the first author) to carrying out this study. All the students were informed that their participation was entirely voluntary. The terms of use agreement for system log data were given to each student for permission to use their data only for the purposes of research aimed at education or learning. Students could accept or decline to indicate their willingness before using the e-book system. All of the participants agreed the terms and were assured that their data would remain strictly confidential.

Each unit had one lecture per week. Each lecture for both classes was delivered in a computer classroom at the participants' university and was taught by the same teacher. That is to say, the teaching and learning content was the same. Students were asked to preview the slides of the lecture using BookRoll before the class. When viewing the slides, they could use a total of 14 available independent features (i.e., reading operations) classified into five categories (Table 1). The participants could click to the next/previous page, search by entering keywords, add or delete bookmarks/markers/memos, or jump to another page using the bookmark or page-change slider on the system via browser. To encourage students at the beginning of each class to access an outline and key information for learning the curriculum, the instructor took 10 minutes to familiarise students' with the overview status of the unit slides using the analysis dashboard of BookRoll, in terms of the number of markers (highlights) for each slide, marker list and text, the number of comments on each slide, memo list and text, and the reading progress/completion of the students.

Table 1
Student operations while reading on BookRoll

Category	Feature	Example
File	• Opening a lecture slide	• A student opens a lecture slide on the system.
	• Leaving the slide being viewed	• A student closes a lecture slide on the system.
	• Searching a keyword	• A student uses the keyword search function on the system.
Page	• Moving to the next page	• A student clicks the next function to the next page.
	• Returning to the previous page	• A student clicks the previous function to the previous page.
	• Jumping to another page using a page-change slider	• A student uses the page slider to jump to another page.
Bookmark	• Attaching a bookmark	• A student uses the bookmark function to attach a bookmark on a particular page.
	• Removing a bookmark	• A student uses the bookmark function to remove a bookmark on a bookmarked page.
	• Jumping to the bookmark page	• A student clicks a bookmarked sign (page index) to the bookmarked page.
Highlight	• Adding a marker	• A student uses the highlight function to add a yellow (unclear) or red (important) marker.
	• Deleting a marker	• A student clicks a yellow or red marker to delete the marker.
Memo	• Writing a memo	• A student uses the memo function to add and write a memo.
	• Removing a memo	• A student deletes a memo through the memo function.
	• Modifying a memo	• A student modifies a memo through the memo function.

Measures

The measures in this study included students' online reading behaviours of previewing lecture slides in the e-book system and their academic achievement. The reading behaviours recorded within the e-book system for the students of each class, in conjunction with their academic achievement in the course, were collected and extracted to form a dataset (two datasets in total).

Student online reading behaviours
The students' actions while reading with regard to the 14 e-book system features were recorded (as students' reading logs) in a database built in a server at the participants' university. This study retrieved the numbers of operations for each participant during the course period for each feature. The number of times each of the 14 e-book system features occurred was counted for analysis, including opening/leaving a lecture slide, searching a keyword, moving to the next/previous page and jumping to another page, attaching/removing a bookmark and jumping to the bookmark page, adding/deleting a marker, and writing/removing/modifying a memo.

Student academic achievement
The final exam scores for the students of this course represented their academic achievement in this study. This final exam was developed based on the lecture slides to evaluate achievement in the course. The exam items were formulated by professors in the College of Information and Electrical Engineering, and the appropriateness of this exam was confirmed by an experienced professor who is an expert in student information and computer education. This exam included content taught in the seven units, and consisted of 35 multiple-choice questions, each with four answer choices, thus, there were 5 questions for each unit.

For example the following question was used to assess concepts regarding operating systems and task managers:

Which of the following is not the part of the operating system that is responsible for secondary memory management?
  (a) Manage unused space
  (b) Provide a mechanism for synchronisation between processes
  (c) Allocate space to needed files
  (d) Disk scheduling

Students' answers were scored (2 points for a correct answer) to calculate the exam scores, with the scores of this exam totaling 70 points. The students were required to finish this test in 30 minutes. Reliability was established via evaluation of exam results from 103 students of the same age (18 to 19 years old): a different group to the 100 research participants. The Kuder-Richardson reliability index (KR20) was .83.

Course implementation

The course of this study was conducted over nine weekly classes in a computer classroom, including: 1-week training, 7-week experimental course, and 1-week final exam. Before the experimental course began, the students received brief training (15 minutes) on how to view lecture slides on the e-book system, including the use of bookmarks, highlighting, and memo functions. The procedure for the experimental course is illustrated in Figure 1. Starting from the second week, the experimental course lasted 7 weeks (one unit per week in two 50-minute classroom periods). For each unit, the teacher uploaded the lecture slides to BookRoll, and the students were asked to preview the slides before class. The participants read materials using different e-book functions in their own ways. Students were free to use the e-book system after class. The final exam was administered in the last week (30 minutes).
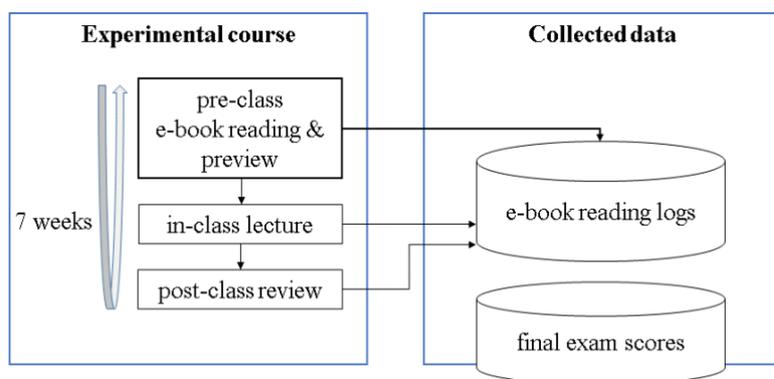

Figure 1. Procedure of the course and collected data from the experiment

Prediction implementation

The prediction procedure included data preprocessing, and classification, and evaluation.

Data preprocessing
Students' reading logs and their achievement test scores were collected. This study extracted the 14 features from the e-book system database for the two classes' datasets. Students' achievement prediction performance is correlated to grading standards/policies, they can be categorised into traditional, lenient, and stringent grading scales to define the labels of student learning performance (Elikai & Schuhmann, 2010; Huang et al., 2020). To better understand prediction performance under different grading policies, this study also used traditional, lenient, moderate, and stringent grading policies. Students who did not reach 60% of the maximum score, and the bottom 20%, 50%, and 80% of students, were labeled as lower-score groups, respectively. That is, students who got a mark of 60% or more on the final exam, as well as the remaining top 80%, 50%, and 20% of all students, were labeled as higher-score groups, respectively.

Classification and evaluation

This study employed six classification algorithms, implemented by Python 3.5, including LR (with L2-norm penalty), GaussianNB, SVC (with linear kernel), DT (with Gini criterion), RF (with 100 estimators and mean square error criterion), and NN (with 100 layers, ReLU activation, and Adam solver), to classify the academic achievement of students based on their reading logs. To evaluate the prediction performance of student achievement classification, this study used 3-fold cross-validation, suggested by Kyoto University (2018) for a relatively small sample size in an educational and learning activity. Of the three subsamples, one subsample was retained as validation data for testing the prediction model, while the remaining two subsamples were used as training data. The values of the 14 features for each subsample were normalised using *z*-scores to transform the range of these values into a smaller and more specific range, as the original range was too wide (e.g., the number of times of moving to the next page was much higher than that of removing a memo). The process of modeling included training by using the classification algorithms to predict the achievement labels from the e-book features and testing the prediction models for the two datasets in this study. In order to better evaluate the prediction performance, various evaluation metrics were applied in this study. This study used four evaluation metrics, including accuracy, precision, recall, and the area under the ROC curve (AUC), to measure the prediction performance of the six algorithms. Evaluation of the models was calculated by taking the average of 3-fold cross-validation that had been run 500 times with partition fold selected randomly for each run. In addition, to understand the key factors that could affect the prediction performance, this study used the Spearman's rank correlation to investigate the relationship between the features of e-book system (i.e., student online reading behaviours) and student academic achievement (i.e., their final exam scores), as the Spearman's rank correlation is less sensitive and still robust to strong outliers (de Winter et al., 2016).

Results

Data from three students who were absent for more than three lectures and did not use the e-book system was unavailable and excluded from the analyses. Thus, there were 97 (52 and 45, respectively, in the two classes) students entering the following analyses. The two classes' original logs had 51,993 and 45,195 records, with averages of 111 and 112 records per student per week, respectively.

Table 2 displays the means and standard deviations of the number of occurrences of each extracted feature in both datasets. Table 3 provides a brief description of students' academic achievement in each dataset under different grading policies: traditional, lenient, moderate, and stringent grading. These data were evenly distributed, and the distribution of the students' scores was close to normal. Approximately one-third of the students in this course did not reach the traditional standard (60% of the total score). The rest of this section answers the two research questions according to the proposed hypotheses.

Table 2
Descriptive statistics of student online reading behaviours in BookRoll for the two datasets

Category	Feature	Dataset 1[a]		Dataset 2[b]	
		M	*SD*	*M*	*SD*
File	Opening a lecture slide	18.31	13.41	19.02	16.49
	Leaving the slide being viewed	9.69	9.55	7.49	8.71
	Searching a keyword	1.31	3.35	2.89	11.69
Page	Moving to the next page	485.54	344.65	535.27	455.21
	Returning to the previous page	201.92	195.86	204.02	220.39
	Jumping to another page	26.04	40.31	22.31	30.00
Bookmark	Attaching a bookmark	10.48	15.40	5.93	12.66
	Removing a bookmark	1.60	3.46	0.69	1.18
	Jumping to the bookmark page	38.23	69.12	13.20	31.11
Highlight	Adding a marker	83.94	95.97	67.87	82.24
	Deleting a marker	11.08	20.95	7.91	12.45
Memo	Writing a memo	6.46	14.11	3.71	9.72
	Removing a memo	0.40	1.01	0.44	1.49
	Modifying a memo	8.40	28.35	3.84	10.69

Note. [a]$N_1 = 52$. [b]$N_2 = 45$.

Table 3
Description of student academic achievement for the two datasets

Dataset	Grading policy	n Lower-score student	n Higher-score student	M Lower-score student	M Higher-score student	SD Lower-score student	SD Higher-score student	95% CI Lower-score student	95% CI Higher-score student
1	Traditional	15	37	34.53	55.24	3.34	8.28	32.69–36.38	52.48–58.00
	Lenient	12	40	33.33	54.05	2.46	9.02	31.77–34.90	51.17–56.93
	Moderate	30	22	40.67	61.00	7.21	4.61	37.98–43.36	58.96–63.04
	Stringent	43	9	45.86	65.56	10.07	2.60	42.76–48.96	63.55–67.56
	Overall			49.27		11.88		45.96–52.58	
2	Traditional	10	35	34.60	53.03	4.99	7.58	31.03–38.17	50.42–55.63
	Lenient	10	35	34.60	53.03	4.99	7.58	31.03–38.17	50.42–55.63
	Moderate	24	21	41.00	58.00	6.78	5.14	38.14–43.86	55.66–60.34
	Stringent	37	8	45.78	63.50	8.64	3.16	42.90–48.66	60.86–66.14
	Overall			48.93		10.47		45.79–52.08	

Note. Maximum possible score of student academic achievement = 70

Machine learning prediction of student academic achievement

Table 4 shows the prediction performance regarding the academic achievement of the six classification algorithms. To answer RQ1: "In a university course supported by an e-book system, to what extent can the academic achievement of students be predicted based on their reading logs by using machine learning classification algorithms?" the prediction performance is presented by evaluation metrics (accuracy, precision, recall, and AUC) in the order of traditional, lenient, moderate, and stringent grading policy. These values ranged from 0 to 1. For instance, a AUC value dropping close to .5 indicates that its classification was similar to random guesswork. Classification with a higher value implies better prediction performance.

According to the results, these models achieved fairly high prediction performance. All classification algorithms (LR, GaussianNB, SVC, DT, RF, and NN) performed well under both traditional and lenient grading policies, ranging from .71 to 1.00 (from moderate to perfect predictions) in addition to those scores with the AUC metric. The results support H1.1 to H1.6, which propose that the LR, GaussianNB, SVC, DT, RF, and NN classifiers perform well based on university students' online reading logs. It can be found that the algorithms such as LR, SVC, DT, RF, and NN were mostly good under different grading policies with the accuracy metric. In addition to GaussianNB, all classification algorithms (LR, SVC, DT, RF, and NN) performed quite well under traditional and lenient grading policies with precision (.82 – .90), while LR, GaussianNB, DT, RF, and NN obtained very good prediction performance with the recall metric (.81 – 1.00) among the two datasets. Moreover, LR, SVC, and NN measured by the AUC also performed well under the moderate grading policy for both datasets (.76 – .85), indicating there was a 76% to 85% chance that the models would be able to distinguish between higher- and lower-group students. Taken together, these results suggest that LR (*M* = .80; range: .65 – .90) and NN (*M* = .80; range: .67 – .89) had the best average predictions with various evaluation metrics under different grading policies, except in the case of stringent grading for predicting student academic achievement.

Table 4
Prediction performance of academic achievement under different grading policies for the two datasets

| | | Evaluation metric | | | | | | | | | | | | | | | | |
| | | Accuracy | | | | Precision | | | | Recall | | | | AUC | | | |
DS	Classifier	T	L	M	S	T	L	M	S	T	L	M	S	T	L	M	S
1	LR	.80	.82	.85	.80	.90	.88	.82	.42	.81	.88	.83	.37	.79	.74	.85	.63
	GaussianNB	.71	.77	.78	.78	.71	.77	.68	.37	1.00	1.00	.93	.33	.50	.50	.80	.60
	SVC	.76	.81	.83	.79	.89	.89	.82	.38	.76	.86	.77	.33	.76	.75	.82	.61
	DT	.74	.78	.81	.78	.82	.86	.79	.37	.81	.86	.76	.43	.69	.69	.80	.64
	RF	.79	.80	.85	.82	.87	.87	.85	.51	.83	.87	.78	.21	.76	.71	.87	.58
	NN	.77	.81	.84	.78	.85	.87	.80	.33	.82	.88	.85	.25	.73	.72	.84	.57
2	LR	.77	.76	.76	.80	.85	.84	.77	.42	.86	.86	.70	.37	.66	.65	.76	.63
	GaussianNB	.78	.78	.47	.72	.78	.78	.47	.33	1.00	1.00	.99	.44	.50	.50	.50	.61
	SVC	.73	.73	.77	.79	.85	.85	.79	.39	.79	.80	.69	.33	.64	.64	.76	.61
	DT	.77	.77	.64	.80	.83	.83	.60	.46	.89	.88	.71	.53	.63	.63	.64	.70
	RF	.76	.76	.72	.84	.82	.82	.71	.56	.89	.89	.69	.37	.60	.60	.72	.65
	NN	.80	.79	.77	.80	.86	.85	.77	.43	.89	.89	.73	.39	.68	.67	.77	.64

Note. DS = dataset; T = traditional grading; L = lenient grading; M = moderate grading; S = stringent grading

Student online reading behaviours affecting the prediction models

To reply to RQ2: "Corresponding to the previous question, what reading behaviours can affect the predictions of student academic achievement?" Spearman correlation analyses were conducted to examine the key factors affecting the prediction performance of classification for academic achievement based on student reading logs. That is, the relationship between the extracted features and students' final exam scores was identified. The resulting correlation coefficients can be found in Table 5. As evidenced by the correlations between the numbers of student operations for the e-book system features (i.e., students' online reading behaviours in the e-book system) and students' academic achievement, the results show that among the five categories situated within the total of 14 features, four categories containing 10 features, including opening/leaving a lecture slide, moving to the next/previous page, jumping to another page, adding/deleting a marker, and writing/removing/modifying a memo, were significantly and positively related to student academic achievement in both datasets, with coefficients ranging from .299 to .775. The results support H2, which states that e-book reading behaviours correlate with academic achievement in the prediction model. Nevertheless, the features of the bookmark category were not correlated to student achievement in dataset 2, and searching keywords was not significantly correlated in either dataset.

Table 5
Correlations between students' online reading behaviours and academic achievement in the two datasets

| | | Coefficient | |
Category	Feature	Dataset 1	Dataset 2
File	Opening a lecture slide	.733***	.648***
	Leaving the slide being viewed	.630**	.299*
	Searching a keyword	.086	.274
Page	Moving to the next page	.760***	.690***
	Returning to the previous page	.764***	.674***
	Jumping to another page	.455***	.521***
Bookmark	Attaching a bookmark	.447***	.111
	Removing a bookmark	.428**	< .001
	Jumping to the bookmark page	.438**	.216
Highlight	Adding a marker	.601***	.312*
	Deleting a marker	.775***	.495***
Memo	Writing a memo	.420**	.364*
	Removing a memo	.388**	.324*
	Modifying a memo	.326*	.413**

Note. $*p < .05$. $**p < .01$. $***p < .001$

Discussion

Machine learning prediction of student academic achievement

This research attempted to use machine learning classification algorithms (including LR, GaussianNB, SVC, DT, RF, and NN) to predict academic achievement among university students in a formal course based on their reading logs from an e-book system, and subsequently explored the critical factors (which will be addressed in the following subsection) that affected predictions. This study took place over a half-semester (9 weeks) of empirical research using e-book system features as inputs and students' final exam scores as outputs in prediction models—labeling lower- and higher-score students as at-risk and non-risk students, respectively—under different grading policies (traditional, lenient, moderate, and stringent). It evaluated the classification/prediction performance with various evaluation metrics (accuracy, precision, recall, and AUC). Being similar in nature, and thus regarded as reliable, good prediction results were obtained from two different classes of the same teacher and course. Based on those results, suitable classification algorithms and grading policies were suggested for performing predictions with different evaluation metrics. In summary, except with regard to the stringent grading policy, GaussianNB was particularly prominent with the recall metric for identifying at-risk students only, while LR and NN had relatively stable prediction performance with different metrics and grading policies for predicting student academic achievement.

Our findings are not in contradiction with those of Waheed et al. (2020), who implemented a deep artificial neural network on a set of unique handcrafted features extracted from virtual learning environment clickstream data to predict at-risk students, with the deep artificial neural network outperforming the baseline SVM model (achieving an accuracy of 78% to 89%). In addition, our result is partly in agreement with Abu Zohair's (2019) finding that the NB and SVM algorithms had achieved 71% and 76% accuracy rates in predicting the performance of university students in a specific course, respectively. However, one classifier he used was SVM with radial kernel. Our results could due to the requirement that students in the course to preview lecture slides and there is a relatively strict grading scale with a reliable instrument. This possibly resulted in a balanced distribution of their exam scores (nearly one-third of students did not reach the traditional pass line of 60 points on the final exam). Students who did not read their materials on the e-book system found it challenging to achieve high marks on the final exam. As such, the classification algorithms could predict results of the higher-scoring group of students more accurately. As researchers (Huang et al., 2020) have suggested, a teacher's grading policy is an essential factor influencing prediction performance. This may lead to poor predictive outcomes due to disparities in the distributions of academic scores among students.

The findings of this study also lend some credence to Huang et al.'s (2020) finding that the AUC was nearly .7 using LR in a university class for a different subject. Furthermore, it is worth noting that the GaussianNB algorithm reached perfect recall scores of 1 (except in the case of the stringent grading policy) in our research, indicating the labelling of all at-risk students was performed correctly, although GaussianNB was not particularly prominent with other metrics. Precision and recall metrics ignore true negatives that are actually non-risk students. As such, the findings indicate that there very few actual non-risk students were predicted incorrectly as at-risk students using precision (i.e., there were almost no false-alarm predictions), and no actual at-risk students that were not identified using the recall metric (i.e. false negatives, which are usually less tolerable in the field of education). Recall should be favoured when false negatives are less tolerated (Özler, 2019). Therefore, in considering the findings with regard to the evaluation metrics—recall and AUC—educators could prioritise employing GaussianNB, LR, or NN when aiming to identify at-risk students in an e-book supported course or learning activity.

Overall, the applied classification algorithms constructed good academic achievement prediction models based on students' reading logs. However, it is worth noting that there is no universal classifier that can perform effectively with identical dataset characteristics in different use case scenarios (Asif et al., 2017; Lu et al., 2021). This is particularly evident through prediction performance achieving a mere .79 (a moderate prediction) with the AUC, except when using the moderate grading policy in this study.

Student online reading behaviours affecting the prediction models

On the whole, the significant correlation coefficients for measuring the relationship between the features and labels were mostly moderate-to-high, and it is therefore possible that the prediction performance in

both datasets had fairly high values with various evaluation metrics. The findings indicate that students' online behaviours of previewing lecture slides in the e-book system were closely and positively correlated to their academic achievement. The most critical features affecting the prediction of student academic achievement included: opening the e-book (i.e., starting to read materials), moving to the next or returning to the previous page (i.e., studying the slides), jumping to another page (i.e., jumps in pages while reading), and deleting a marker. The finding with regard to returning to the previous page confirms Yin et al.'s (2019) finding that student reading behaviour of *backtrack* reading had a significant positive influence on learning effectiveness. Although the bookmark category's features were highly correlated to student achievement in dataset 1, the reasons why they were not significant in dataset 2 may be explained by their relatively small numbers of occurrence, as seen in Table 2 (M = 5.93 and 0.69). Similar results can be seen in Table 2 with both datasets for searching a keyword (M = 1.31 in dataset 1, and 2.89 in dataset 2). As a feature which was rarely used, this may be partially responsible for the no significant correlation between this feature and academic achievement.

The features that had positive and significant correlation with academic achievement may have something to do with some constructive reading behaviours. For instance, for the features of writing and modifying memos, students had to elaborate the course material content. This reading behaviour was akin to note-taking, which can increase students' comprehension of a lecture (Chen et al., 2012). Returning to the previous page and jumping to another page refers to the seeking of concepts, while adding or deleting markers (highlights) means focusing on specific concepts or keywords that the students identified with markers. Yang et al. (2021) suggested that students who adopt reflective learning and have a tendency to do backtrack reading and delete annotations from digital materials can achieve better learning outcomes. Moreover, bookmarks act as reminders for reflection. As such, the students made it easy to flash back to those important slides. However, the features of bookmarking were only significantly correlated to student academic achievement in dataset 1. Future studies could encourage students to utilise the bookmark function in an e-book learning activity and investigate whether positive correlations exist between student achievement and these features that affected the prediction model.

Implications for further research and pedagogical practice

This study extends prior research on the use of machine learning classification algorithms in an e-book supported university course, obtaining good prediction performance of academic achievement. This research also contributes to proving suitable prediction methods for predicting student academic performance based on their digital reading behaviours in a higher education setting. This could serve to predict at-risk students and promote online reading behaviours (in an e-book system) that have been found to be particularly correlated to academic achievement, such as returning to important pages, locating pages with the sliders or bookmarks (for targeting concepts), and utilising highlight and memo functions (for identifying concepts). It is also suggested that future research might build prediction models using machine learning classification methods for teaching material delivery systems or e-book supported courses and apply the models to the courses to provide early predictions of student academic performance (perhaps at the halfway point of the course duration), paying attention to any critical features that may affect the prediction model.

Note that the features analysed in this study were students' online reading behaviours of previewing lecture slides confined to BookRoll. Further studies using different e-book systems are required to validate the findings, since different systems could result in different reading patterns. Furthermore, the output labels in this study were final exam scores of information and technology, hence, caution should be taken in generalising the findings to other assessment measures (e.g., midterm exam or report) or other subjects. Further empirical research with a larger number of participants and datasets from different universities or countries on this topic is required to determine the prediction performance of student achievement as measured by various assessment methods, as well as the correlations between different features and labels.

Acknowledgments

References

Abu Zohair, L. M. (2019). Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*, 27. https://doi.org/10.1186/s41239-019-0160-3

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Using learning analytics to detect off-task reading behaviors in class. In J. Cunningham et al. (Ed.), *Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference* (pp. 471–476). Society for Learning Analytics Research. https://www.solaresearch.org/core/companion-proceedings-of-the-9th-international-learning-analytics-and-knowledge-conference-lak19/

Ashraf, A., Anwer, S., & Khan, M. G. (2018). A comparative study of predicting student's performance by use of data mining techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences*, *44*(1), 122–136. https://www.asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/4170

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, *113*, 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

Bekele, R., & McPherson, M. (2011). A Bayesian performance prediction model for mathematics education: A prototypical approach for effective group composition. *British Journal of Educational Technology*, *42*(3), 395–416. https://doi.org/10.1111/j.1467-8535.2009.01042.x

Chen, C.-H., Chen, Y.-X., Chow, Y.-H., & Pan, S.-H. (2020). Investigating and predicting the usability of an e-book system for university students: The role of prior knowledge. *Communications in Computer and Information Science, 1227*, 333–342. https://doi.org/10.1007/978-981-15-6113-9_38

Chen, C.-H., Chiu, C.-H., & Wu, C.-Y. (2012). Effects of shared note-taking and questioning review in elementary school computer classes. *Contemporary Educational Research Quarterly*, *20*(2), 47–91. https://doi.org/10.6151/CERQ.2012.2002.02

Chen, C.-H., & Su, C.-Y. (2019). Using the BookRoll e-book system to promote self-regulated learning, self-efficacy and academic achievement for university students. *Educational Technology & Society*, *22*(4), 33–46. https://www.jstor.org/stable/26910183

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Cheng, G. (2017). Towards an automatic classification system for supporting the development of critical reflective skills in L2 learning. *Australasian Journal of Educational Technology, 33*(4), 1–21. https://doi.org/10.14742/ajet.3029

de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273. https://doi.org/10.1037/met0000079

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. *Proceedings of the 7th International Learning Analytics and Knowledge Conference* (pp. 188–197). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027447

Dupin-Bryant, P. A. (2004). Pre-entry variables related to retention in online distance education. *The American Journal of Distance Education*, *18*(4), 199–206. https://doi.org/10.1207/s15389286ajde1804_2

Elikai, F., & Schuhmann, P. W. (2010). An examination of the impact of grading policies on students' achievement. *Issues in Accounting Education*, *25*(4), 677–693. https://doi.org/10.2308/iace.2010.25.4.677

Flanagan, B., Majumdar, R., Takii, K., Ocheja, P., Chen, M. A., & Ogata, H. (2020). Identifying student engagement and performance from reading behaviors in open eBook assessment. In H.-J. So., M. M. Rodrigo., J. Mason., & A. Mitrovic. (Eds.), *Proceedings of the 28th International Conference on Computers in Education* (Vol. 1, pp. 235–244). Asia-Pacific Society for Computers in Education.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119. https://doi.org/10.1016/j.ijinfomgt.2019.02.003

Hämäläinen, W., & Vinni, M. (2010). Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. Baker (Eds.), *Handbook of educational data mining* (pp. 57–74). CRC Press. https://doi.org/10.1201/b10274

Heuer, H., & Breiter, A. (2018). Student success prediction and the trade-off between big data and data minimization. In D. Krömker, & U. Schroeder (Eds.), *Proceedings of the 16th e-Learning Conference of the German Computer Society* (pp. 219-230). German Informatics Society. https://dl.gi.de/handle/20.500.12116/21041

Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin, C. J., & Yang, S. J. H. (2020). Predicting students' academic performance by using educational big data and learning analytics: Evaluation of classification methods and learning logs. *Interactive Learning Environments, 28*(2), 206–230. https://doi.org/10.1080/10494820.2019.1636086

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, *61*, 133–145. https://doi.org/10.1016/j.compedu.2012.08.015

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jin, Q., Imbrie, P. K., Lin, J. J. J., & Chen, X. (2011). A multi-outcome hybrid model for predicting student success in engineering. *Proceedings of the 118th ASEE Annual Conference and Exposition* (pp. 22.70.1–12). American Society for Engineering Education. https://peer.asee.org/17352

Kitto, K., Lupton, M., Daavis, K., & Waters, Z. (2017). Designing for student-facing learning analytics. *Australasian Journal of Educational Technology, 33*(5), 152–168. https://doi.org/10.14742/ajet.3607

Kotsiantis, S. B., Patriarcheas, K., & Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, *23*(6), 529–535. https://doi.org/10.1016/j.knosys.2010.03.010

Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. In V. Palade, R. J. Howlett, & L. Jain (Eds.), *Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 267–274). Springer. https://doi.org/10.1007/978-3-540-45226-3_37

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220–232. https://www.jstor.org/stable/26388400

Lu, O. H. T., Huang, A. Y. Q., & Yang, S. J. H. (2021). Impact of teachers' grading policy on the identification of at-risk students in learning analytics. *Computers & Education, 163*, 104109. https://doi.org/10.1016/j.compedu.2020.104109

Luan, H., & Tsai, C.-C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society, 24*(1), 250–266.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, *53*(3), 950–965. https://doi.org/10.1016/j.compedu.2009.05.010

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, *103*, 1–15. https://doi.org/10.1016/j.compedu.2016.09.005

Marbouti, F., Diefes-Dux, H. A., & Strobel, J. (2015). Building course-specific regression-based models to identify at-risk students. *Proceedings of the 122nd ASEE Annual Conference and Exposition* (Paper ID: 11229). American Society for Engineering Education. https://doi.org/10.18260/p.23643

Mayilvaganan, M., & Kalpanadevi, D. (2014). Comparison of classification techniques for predicting the performance of students academic environment. *Proceedings of the 2014 International Conference on Communication and Network Technologies* (pp. 113–118). https://doi.org/10.1109/CNT.2014.7062736

Morris, L.V, Wu, S.-S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education*, *19*(1), 23–36. https://doi.org/10.1207/s15389286ajde1901_3

Murty, M. N., & Raghava, R. (2016). Kernel-based SVM. In S. Zdonik., S. Shekhar., J. Katz., X. Wu., L. Jain., D. Padua., X. Shen., B. Furht., V.S. Subrahmanian., M. Hebert., K. Ikeuchi., B. Siciliano., S. Jajodia., & N. Lee. (Eds.), *SpringerBriefs in computer science: Support vector machines and perceptrons* (pp. 57–67). Springer. https://doi.org/10.1007/978-3-319-41063-0_5

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., & Hirokawa, S. (2017).

Learning analytics for e-book-based educational big data in higher education. In H. Yasuura, C. Kyung, Y. Liu, & Y. Lin (Eds.), *Smart sensors at the IoT frontier* (pp. 327–350). Springer. https://doi.org/10.1007/978-3-319-55345-0_13

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). e-Book-based learning analytics in university education. In H. Ogata, W. Chen, S. C. Kong, & F. Qiu (Eds.), *Proceedings of the 23rd International Conference on Computers in Education* (pp. 401–406). Asia-Pacific Society for Computers in Education. https://apsce.net/download_data.php?filename=upfile/pdf/proceedings-of-icce2015-01.pdf

Özler, H. (2019). Accuracy trap! Pay attention to recall, precision, F-score, AUC. https://medium.com/datadriveninvestor/accuracy-trap-pay-attention-to-recall-precision-f-score-auc-d02f28d3299c

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30–40. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, *143*, 103676. https://doi.org/10.1016/j.compedu.2019.103676

Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: A case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*, 12. https://doi.org/10.1186/s41239-020-00186-2

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior, 104*, 106189. https://doi.org/10.1016/j.chb.2019.106189

Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2019). The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology,51*(2)*,*347–370. https://doi.org/10.1111/bjet.12836

Wen, C.-T., Chang, C.-J., Chang, M.-H., Chiang, S.-H. F., Liu, C.-C., Hwang, F.-K., & Tsai, C.-C. (2018). The learning analytics of model-based learning facilitated by a problem-solving simulation game. *Instructional Science*, *46*(6), 847–867. https://doi.org/10.1007/s11251-018-9461-5

Wu, J.-Y., Hsiao, Y.-C., & Nian, M.-W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments, 28*(1), 65–80. https://doi.org/10.1080/10494820.2018.1515085

Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers & Education*, *43*(4), 345–359. https://doi.org/10.1016/j.compedu.2003.09.005

Xenos, M., Pierrakeas, C., & Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education*, *39*(4), 361–377. https://doi.org/10.1016/S0360-1315(02)00072-6

Yang, A., Chen, I. Y. L., Flanagan, B., & Ogata, H. (2020). Applying key concepts extraction for evaluating the quality of students' highlights on e-Book. In H.-J. So., M. M. Rodrigo., J. Mason., & A. Mitrovic. (Eds.), *Proceedings of the 28th International Conference on Computers in Education* (Vol. 1, pp. 284–288). Asia-Pacific Society for Computers in Education.

Yang, C. C. Y., Chen, I. Y. L., & Ogata, H. (2021). Toward precision education: Educational data mining and learning analytics for identifying students' learning patterns with ebook systems. *Educational Technology & Society, 24*(1), 152–163. https://www.jstor.org/stable/26977864

Yang, S. J. H., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence, 2*, 100008. https://doi.org/10.1016/j.caeai.2021.100008

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human factor approach. *International Journal of Human–Computer Interaction, 35*(4-5), 313–322. https://doi.org/10.1080/10447318.2018.1543077

Corresponding author: Chien-Yuan Su, cysu@mail.nutn.edu.tw