Article

Journal of Educational Computing Research
0(0) 1–26

DOI: 10.1177/07356331211038168
journals.sagepub.com/home/jec
SAGE

An Interpretable Pipeline for Identifying At-Risk Students

Bo Pei[1] and **Wanli Xing**[1]

Abstract

This paper introduces a novel approach to identify at-risk students with a focus on output interpretability through analyzing learning activities at a finer granularity on a weekly basis. Specifically, this approach converts the predicted output from the former weeks into meaningful probabilities to infer the predictions in the current week for maintaining the consecutiveness among learning activities. To demonstrate the efficacy of our model in identifying at-risk students, we compare the weekly AUCs and averaged performance (i.e., accuracy, precision, recall, and f1-score) over each course with the baseline models (i.e., Random Forest, Support Vector Machine, and Decision Tree), respectively. Furthermore, we adopt a Top-*K* metric to examine the number of at-risk students that the model is able to identify with high precision during each week. Finally, the model output is interpreted through a model-agnostic interpretation approach to support instructors to make informed recommendations for students' learning. The experimental results demonstrate the capability and interpretability of our model in identifying at-risk students in online learning settings. In addition to that our work also provides significant implications in building accountable machine learning pipelines that can be used to automatically generated individualized learning interventions while considering fairness between different learning groups.

Keywords

learning analytics, identifying at-risk students, educational data mining, predictive learning model, interpretable model

[1]School of Teaching & Learning, College of Education, University of Florida, Gainesville, United States

Corresponding Author:
Wanli Xing, School of Teaching & Learning, College of Education, University of Florida, Gainesville, FL 32611, United States.
Email: wanli.xing@coe.ufl.edu

Introduction

The increasing popularity of online learning has offered massive opportunities and flexibilities to students for accessing quality learning experiences since the last decade. Serving as an important alternative learning format to face-to-face learning, learning online has become ubiquitous during the COVID-19 pandemic (Leite et al., 2021). This learning format makes it possible for students that are geographically dispersed to learn together and provides instructors with detailed digital learning traces that are used to examine students' learning efforts. Approaches (Coussement et al., 2020; Hooshyar et al., 2019) have been investigated extensively to reveal students' learning patterns underlying the learning traces and analyze how the differences in the learning patterns correlate with the differences in learning performances. The analytic results allow researchers, educators, and policy-makers to make informed decisions in terms of predicting learning performance, identifying at-risk students and providing individualized interventions (Coussement et al., 2020; Hidalgo et al., 2021; Zhang et al., 2020).

Currently, the fundamental dilemmas that undermine the practices of online learning are the high dropout rate and the low learning achievements compared with that in the face-to-face learning settings (Aljohani et al., 2019; Hassan et al., 2019; Rovira et al., 2017; Xing, Tang, et al., 2019). This is in part due to that the students in online learning environments are required to exert more effort to maintain the same level of achievements as in face-to-face learning settings. Multiple studies (Carvalho et al., 2020; Onah et al., 2014; Tang et al., 2019) indicated that instructors play a crucial role in preventing students from dropping out of online courses by providing at-risk students with timely individualized interventions to encourage them to stay in the courses. However, in online learning settings, it is nearly impossible for instructors to identify struggling students, let alone provide timely effective interventions because of the large number of participants and little face-to-face communication opportunities (Gunnarsson & Alterman, 2012; Xing, Pei, et al., 2019).

Over the years, a considerable amount of effort has been devoted to finding ways to address these issues by incorporating machine learning and deep learning approaches into learning analytics (Chen et al., 2019; Waheed et al., 2020; Xing & Du, 2019). However, those models have been demonstrated with mixed performance in terms of both performance and interpretability while applied in real settings. Generally, this is caused by inherent mechanisms while building the models: On the one hand, most machine learning models (i.e., Logistic Regression, Decision Tree) used currently were built without considering the associations among the learning behaviors. Moreover, the feature extraction approaches (Liang et al., 2016) that are commonly used by the machine learning models are not built for dealing with temporal-related information, which also impacts the capabilities of these models in capturing the time-sensitive information in the dataset; On the other hand, while the approaches (Fei & Yeung, 2015;

Tan et al., 2018) such as Recurrent Neural Networks, Long-Short Term Memory Networks are built for dealing with the sequential information, they always have to maintain complex network structures to model the associations among the high dimensional data, which may result in high computational complexity and low output interpretability.

Here, we developed a framework to model and analyze the students' learning behaviors to predict at-risk students with high accuracy and generate interpretable output for instructors to make informed decisions in support of personalized learning. The model achieves early predictions of at-risk students by incorporating the insights drawn from the previous weeks into the predictions in the current week. By considering more the associations among the learning activities, the prediction curves of the students' dropout probabilities could be much smoother across weeks. Apart from this, this weekly inferring strategy makes it possible to train the models without the requirements of the learning data from previous similar courses, which further increases the generalizability of the models. Finally, to provide accountable individualized interventions for students, we also employed the explanatory mechanisms for interpreting the predicted results to make the decision-making process transparent to instructors rather than treat them as "black-boxes".

The rest of the work goes as follows: In "Literature Review" section, we review the previous studies on the practice of identifying at-risk students in online learning environments. In "Methodology" section, we introduce the problem definition and model construction processes. In "Dataset Description and Feature Preparation" section, we describe the dataset that will be used in the paper and conduct the related feature extraction analysis for the models that will be involved in the paper. In "Experiment and Result Analysis" section, we experimentally compare and evaluate the models' performance in identifying at-risk students under different metrics as well as interpret the output for providing actionable interventions. In "Discussion" section and "Conclusion" section, we discuss our findings and make conclusions about our work as well as recommendations for future research directions.

Literature Review

Identifying at-risk students accurately at an early stage is one of the major focuses in the practices of online learning analytics. With the at-risk students identified, instructors can conduct various subsequent activities (e.g. individualized interventions, adapting instructional strategies) to retain these students in the course and help them to learn as much as they can. As such, the literature review in this paper mainly focuses on the approaches that have been investigated to predict at-risk students precisely.

Machine learning algorithms have been extensively investigated for providing accountable results for predicting at-risk students in online learning environments. For example, in He et al. (2015), the authors first explored the feasibility of applying machine learning approaches to predict the probability that students drop out of an online course based on their dynamic learning activities. They adopted the Logistic Regression model to predict students' dropout probabilities at the end of each week. The study indicated that the contributions of the learning activities to the prediction result can be drawn from the model's coefficients. With an understanding of the importance of each feature to the predicted result, instructors can provide more specific recommendations to improve the students' performance. In this research, the curve representing the predicted dropout probabilities was smoothed by calibrating the predicted results without considering the underlying learning associations across weeks. Moreover, the model was only evaluated based on the AUC metric, which could not provide information about a specific student's learning states.

In another study, Hlosta et al. (2017) proposed a selfLearner algorithm that built machine learning models based on the learning data generated in the current course to improve the generalizability of the models. The authors in this research specifically examined the associations between the students' first assignment performance and their dropout probabilities in the course. They found that about 94% of students who did not submit the first assignment would withdraw from the course, and 80% of the students who submitted but failed the assignment would drop out later. As such, the proposed selfLearner model was explicitly used to predict students' performance on the first assignment, especially whether they would submit the assignment or not, to provide information about the students with probabilities to drop out. Apart from evaluating the selfLearner model with the traditional metrics such as Accuracy, Precision, Recall, F1-measure, and AUC, the study also adopted the Top-K precision and Top-K recall metrics to evaluate the model performance in identifying the top K at-risk students. In this research, the SelfLearner model inferred the students' dropout probabilities mainly based on students' learning activities before submitting the first assignment, which could not be generated to other courses whose first assignments could not be an indicator for students' dropout probabilities.

Some studies considered the temporal information among the learning activities while building machine learning models but at a relatively coarse granularity. For example, Hlosta et al. (2017) leveraged the average/median of the clickstreams, the normalized number of materials visited per day, and the number of active days as features to build models. Although the authors considered the sequential information by calculating the daily learning activities, the averaged/median and normalized alternatives of students' actual learning behaviors largely undermine the associations between the behaviors performed across different timestamps. Adnan et al. (2021) calculated the sum and average of the clickstreams at the stages of 20%, 40%, 60%, 80%, and 100%

according to the progress of the course, and built models to predict students' learning performance during each stage. Similarly, in this study, the authors did not consider the temporal associations among the learning behaviors during each stage.

Recently, increasing attention has been paid to the applications of deep learning approaches in analyzing students' online learning performance due to the capabilities of dealing with sequential data. Specifically, the approaches such as Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network (Fei & Yeung, 2015; Tan et al., 2018; Xiong et al., 2019) have been applied to explore the sequential correlations between students' learning behaviors and build models to predict the learning behaviors that are more likely to be conducted in future and students' learning performance. Specifically, Tan et al. (2018) proposed a multi-path learning-based scheme to track students' learning behaviors for predicting the probabilities of attrition. This approach employed a neural network structure to keep track of the trends of students' attrition. Similarly, Qiao and Hu (2020) implemented a joint neural network model to analyze both user demographic information and learning behaviors for predicting at-risk students. The experimental results demonstrated the proposed model outperformed the commonly used baseline models in identifying at-risk students. Although these models have a good performance in identifying the at-risk students, they have to maintain a complex network structure with huge computational costs during the model training process. Moreover, the network structure makes it difficult to generate the interpretable output that can be generalized to a wider environment and used for other purposes.

Building upon the above studies, in this study, we propose a novel approach that can support timely, accurate, and interpretable predictions by focusing on the consecutiveness of the learning behaviors. Specifically, the predicted results from the previous weeks were converted to meaningful insights that can be used to make predictions in the current week to ensure the prediction curve is more in line with students' actual learning patterns. Apart from that, we also adopt the metrics such as Top-K precisions and Top-K recalls to evaluate the model's performance in identifying the most at-risk students to facilitate the resource allocation processes. Finally, we further provided actionable interpretations to help the instructors have a good understanding of the process that how the algorithms made the corresponding decisions and based that to provide more appropriate interventions for student learning.

The main contributions of this study are as follows:

1. A novel framework is proposed to investigate the consecutiveness of prediction curves in identifying at-risk students;
2. Comprehensive experiments are conducted to evaluate the performance of the proposed model against the baseline models using different metrics;

3. Our proposed model has high generalizability and can be easily adopted in other similar environments working on different online courses.

Methodology

Problem Definition

We explore a model that predicts at-risk students weekly by incorporating the predicted probabilities in the previous weeks in the prediction process during the current week. In this paper, we convert the prediction problem into a supervised binary classification problem - classifying whether a student will drop out (1) or stay (0) in the courses in the following week. During the classification process, the probabilities of students' dropout will be calculated by the classifier. To make the prediction results meaningful, we set the threshold as 0.6 to indicate whether a student will drop out or not (He et al., 2015). As such, a larger probability indicates a more urgent intervention is needed from the instructors.

Model Setup: In this section, we focus on how to formulate the problem of predicting at-risk students mathematically. To present the proposed model precisely, we first introduce the following necessary notations. Letting S denote a set of learners, we define the learner's learning activities as follows:

Learning Activity: Let $X^{T \times N \times D}$ be an activity tensor, each element $X_{t \times i \times d}$ represents the d^{th} activity by student $s_i \epsilon S$ at time t indicating a time interval of a week. And T represents the total number of weeks, N is the number of students in that week and D is the count of different activity types. The activity space includes all the activities (e.g. viewing forum, checking the learning resources) that the student performed in the learning system. We use a D-dimensional vector $X^t(i) = [X_{t,i,0}, X_{t,i,1}, \ldots, X_{t,i,d-1}]^T$ to represent all the activities performed by student s_i during week t.

Learning State: $Y \in \{0,1\}^{T \times N}$ denotes a learner's actual learning state (i.e., drop out or not) in a specific target time window $[T, T+1]$, and $y_{t,i}$ denotes the predicted state of student s_i in the window of $[T, T+1]$.

Static Attributes: $A^{T \times M}$ denotes a learner's static attributes (e.g. gender, age) that will not change in a short period, where T is the number of weeks and M is the total number of static attributes. For example, for the gender attribute, we have $A^0(i) = A^1(i) = \ldots = A^{(T-1)}(i)$, which means the student i's gender is not changing across weeks during the course.

In this paper, we combine the static attributes and learning activities together as the predictors of students' learning states during each time interval. So, the attributes $X = [A^{T \times M}, X^{T \times N \times D}] \in R^{T \times N \times (M+D)}$ is a tensor in which $X_{t,i,j}$ represents the jth attribute of student s_i at time t. Our goal in this paper is to find a reasonable mapping rule from the attribute space to the student states space $R(\cdot) : X \rightarrow Y$. Then, we apply the mapping R(.) on the aggregated attributes until the current timestamp to predict students' future states (i.e., whether dropout or not). The dropout probability of student i at the time stamp $t+1$ can be denoted as:

$$p^{t+1}(y_i = 1|X_i^t) \tag{1}$$

Model Construction

In the section above, we introduced the definition of the at-risk prediction problem as well as the involved parameters in formulating the problem mathematically. According to Figure 1, the predicted probability of a student's dropout in the current week $P(w_t)$ depends on that from the previous week $P(w_{t-1})$ and students' learning activities during the current week. In this model, we capture the correlations between the student activities and latent learning states based on the sigmoid function $f(\cdot)$, considering its capabilities of providing interpretable results.

$$f(x) = \frac{1}{1+e^{-z}} \tag{2}$$

Specifically, the probabilities of dropout based on the combined features Z_i^t at time t can be modeled by

$$y_i^t = f(w^T \cdot Z_i^t + b) \tag{3}$$

where, $w \in R^{m+d}$, $b \in R^n$ are the parameters for the observed learning activities. And Z_i^t is the combination of student i's demographic information $A^{T\times M}$ and aggregated learning activities $X^{T\times N\times D}$ till time t.

Based on the model and parameters defined above, we define the objective function as follows:

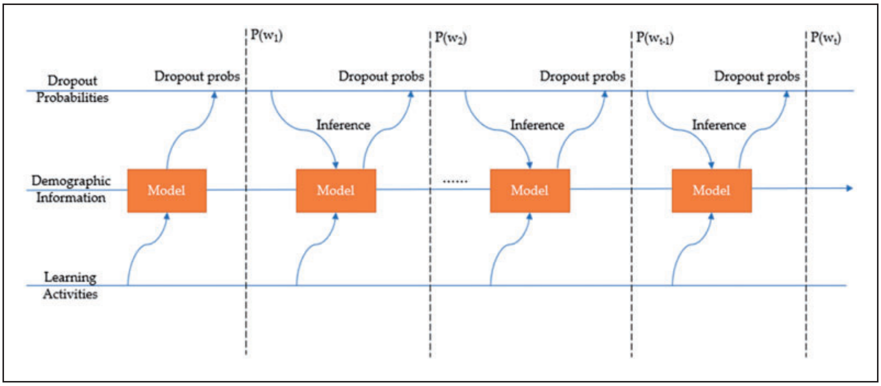

Figure 1. The Schematic Overview of the Proposed Framework.

$$O(\Theta) = \sum_{i=0}^{N-1} \sum_{t=0}^{T-1} \|y^t(i) - y^t(i)^*\|^2 + \lambda_o \sum_{i=0}^{N-1} \sum_{t=1}^{T-1} \|y^t(i) - y^{(t-1)}(i)\|^2 + \lambda_w \|w\|^2 + \lambda_c \sum_{i=0}^{N-1} \sum_{t=0}^{T-1} \|y^t(i) - 0.6\|^2 \tag{4}$$

where, $y^t(i)^*$ is the actual state of the current time window; λ_o is a parameter that controls the differences between the predicted values of the previous time window and the predicted values of the current time window; λ_w is a parameter that controls the regularization value; λ_c is a parameter that controls the distances between the predicted value and the decision boundary.

The learning process of the proposed model is to find the most appropriate parameters of $\Theta = \{\lambda_o, \lambda_w, w, \lambda_c\}$ to minimize the objective function Equation 3 and estimate the latent activities based on historical behaviors. The objective function is convex and has a closed form and its parameters can be learned by optimization methods like the gradient descent method and its variants e.g., Stochastic gradient descent (SGD).

Taken together, this model has two sets of parameters: model parameters Θ and latent learning status y_i^t. In this paper, these parameters were estimated

Table 1. Model Learning Procedure.

Input: total iterations L;
The learning rate in the E Step ρ_e;
The learning rate in the M Step ρ_m;
Output: The learning parameters $\Theta = \{\lambda_o, \lambda_w, w, \lambda_c\}$;
The latent learning status Y;
Initialize the model parameters $\Theta = \{\lambda_o, \lambda_w, w, \lambda_c\}$;
Set the initial latent learning status Y as the actual learning status;
for l=1 to L do
E Step: # fix Θ, update Y;
Compute gradient $\nabla Y_{t,i}$;
Update $Y_{t+1,i} \leftarrow Y_{t,i} + \rho_Y \nabla Y_{t,i}$;
M Step: # fix Y, update Θ;
for n=1 to N do
Calculate the gradient of all parameters and update according to
X, Y in the time span [T, $T+l$];
Update $w \leftarrow w + \rho_w \nabla w$
Update $\lambda_w \leftarrow \lambda_w + \rho_{\lambda_w} \nabla \lambda_w$
Update $\lambda_o \leftarrow \lambda_o + \rho_{\lambda_o} \nabla \lambda_o$
Update $\lambda_c \leftarrow \lambda_c + \rho_{\lambda_c} \nabla \lambda_c$
end
end
return Θ and Y;

based on the Expectation-Maximum (EM) approach, and the detailed process is described in Table 1. The EM algorithm (Mubarak et al., 2020) is an approach that is widely used in the machine learning field for density estimation in the presence of latent variables. This is an iterative approach that can be divided into two separated stages: The E-step attempts to estimate the latent or missing variables; The M-step uses the estimated variables to optimize the parameters to best explain the data.

Dataset Description and Feature Preparation

To demonstrate the validity of the proposed method, we used the OULAD (see https://analyse.kmi.open.ac.uk/open_dataset) dataset collected by Open University (Kuzilek et al., 2017). Multiple studies (Adnan et al., 2021; Aljohani et al., 2019; Hlosta et al., 2018, 2017; Qiao & Hu, 2020) have been conducted using this dataset as a baseline to verify the effectiveness of their methods in predicting student learning performance. In this study, we selected four courses purposely from the dataset, where two of them (i.e., AAA_2013J and AAA_2014J) were in the social science field and the other two (i.e., CCC_2014B and CCC_2014J) were in the STEM field. The reason for selecting these four courses is to build a predictive model with high generalizability by increasing the diversity of the dataset to examine the different learning patterns of students from different fields. In these courses, all the information that can be used to identify the courses and students was anonymized. The detailed information about the features we used to build the model is presented in Table 2.

Figure 2 compares the differences in the distributions on the highest educational levels, number of dropout students, gender, and final performance of the students from the two fields. From the figure, we find that there is a similar distribution on the highest educational levels of students from the social science and stem courses. But large differences are found in the distributions of student genders in the social science and stem courses. Specifically, a significantly large number of male students enrolled in the courses from the stem field. And the students from the stem-related courses are found more likely to drop out of the courses during their learning processes. Moreover, the final performance distributions also indicate that the number of students withdrawn from the stem course is the highest, while in the social science field, the highest number of students is of those who passed the courses.

higher than any other performance category, which further demonstrates a higher probability of dropout in the stem course.

a. The distribution of students' educational levels in different subjects
b. The distributions of gender in different subjects
c. The number of dropout students in different subjects
d. The distribution of students' final performance in different subjects

Table 2. Data Features Used for Building the Model.

Feature type	Examples	Descriptions
Static data	Age	Student demographic information
	Gender	
	Highest educational level	
	Geographical region	
	Number of previous attempts	
	Disability	
Dynamic data	Sum of Resources clicks	Clicks on pdf resources such as books
	Sum of Data_plus clicks	Clicks on the additional information such as videos, audios, sites, etc.
	Sum of Forum clicks	Clicks in the discussion forums
	Sum of Glossary clicks	Clicks on the basic glossary related to contents of course
	Sum of Homepage clicks	Clicks on the course homepage
	Sum of ou_Collaborate clicks	Clicks on the online video discussions
	Sum of ou_Content clicks	Clicks on the content of the assignment
	Sum of the Page clicks	Clicks on the information related to the course
	Sum of Subpage clicks	Clicks on other sites enabled in the course
	Sum of Quiz clicks	Clicks on the course quiz
	Sum of URL clicks	Clicks on the links to audio/video contents
	Number clicks of different resources	The sum of clicks related to different resources
	Minimum clicks	The minimum clicks student performs during the week
	Maximum clicks	The maximum clicks student performs during the week
	Mean of the total clicks	The mean of total clicks

Figure 3 depicts students' averaged resource clickstreams and counts in both the social science and stem courses, respectively. It is obvious that the averaged clickstreams from the stem courses have several obvious peaks in certain periods during the semester, while as for the students from the social science courses, the averaged clickstreams present less abrupt changes during the semester. However, as for the averaged types of resources visited by the students, a similar pattern is found between these two subjects.

Figure 4 shows the number of retaining students in these two courses at the end of each week. From the figure, it can be found that there is a difference in student dropout rates between the Social science and stem courses. In the social science courses, the students maintain almost the same value across weeks during the semester. While in the stem courses, there is a higher dropout rate

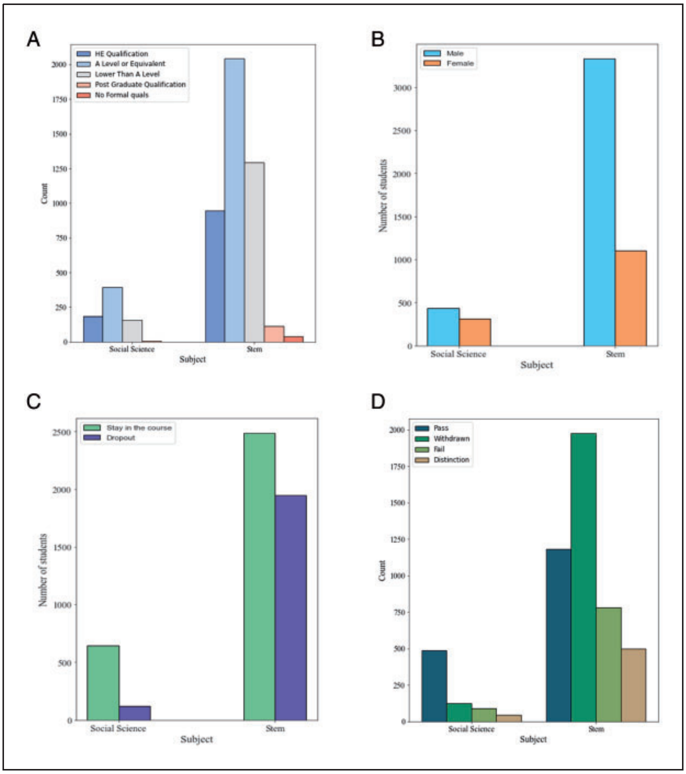

Figure 2. The Different Distributions of Students From Different Subjects. A: The distribution of students' educational levels in different subjects, B: The distributions of gender in different subjects, C: The number of dropout students in different subjects, D: The distribution of students' final performance in different subjects.

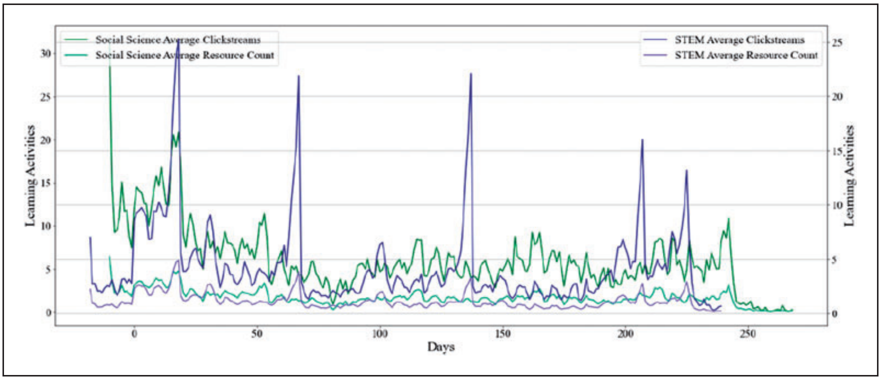

Figure 3. The Patterns of Averaged Clickstreams Students Conducted and Types of Resources Visited Each Day in the Social Science and Stem Courses.

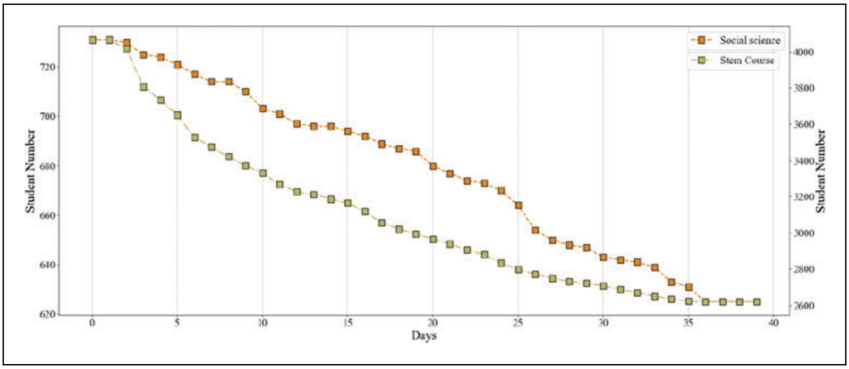

Figure 4. The Comparison of Weekly Dropout of Students in the Social Science and Stem Courses.

compared with that of the social science courses in the first several weeks; and during the several weeks in the middle of the semester, the students from these two subjects share almost the same dropout rate; but in the last several weeks, there seems to be no dropout occurs in the stem courses.

Experiment and Result Analysis

In this section, we conduct experiments to compare the performance of the proposed algorithm with that of the baseline models. Particularly, we first introduce the configurations of our model, the characteristics of the baseline models, the metrics for evaluating the model performance, and the approaches for interpreting the output. Then, the comparison results of the models are presented and analyzed in detail.

Experimental Setup

Model Configurations. We utilized python libraries scikit-learn (see https://scikit-learn.org/stable/) to build and train our proposed model in this paper. Regarding the tunable parameters $\Theta = \{\lambda_o, \lambda_w, \lambda_c\}$ in the model, we found the best combination (i.e., $\lambda_o = 0.5, \lambda_w = 0.01, \lambda_c = 0.05$) using 8 fold cross-validation in our environment. And all the experiments were conducted on an x64 machine with a 2.60 GHz Intel Core i7 CPU and 16.0 GB RAM.

Baseline Models. To demonstrate the effectiveness of the proposed model, in this section, we introduce several machine learning models that are commonly used as baseline models in the literature.

- **Random Forest (RF)** has been demonstrated with a good performance in dropout prediction and is frequently deployed in the industry (Tan et al., 2018). During the prediction process, multiple decision trees are constructed on random attributes to produce generalizable outcomes, which greatly avoids the potentials of overfitting during the training process (Jayaprakash et al., 2020).
- **Support Vector Machine (SVM)** separates different classes based on learned hyperplanes that maximize the margins between different categories (Marbouti et al., 2016). The performance of the SVM models is only sensitive to the samples that are close to the determination surface, so these models usually present good generalizability. In educational settings, Linear SVM is the model used most frequently to predict student learning performance (Naicker et al., 2020).
- **Decision Tree (DT)** separates data based on the partitioning mechanisms. Specifically, the features of the dataset are compared with a predefined criterion until all the samples are partitioned into separated classes. The purity of the separated classes is generally assessed by the Gini Index (Chen et al., 2019).

Evaluation Metrics. We adopted various metrics to measure the involved models' performance in identifying at-risk students. These metrics include the commonly used approaches for evaluating the performance of machine learning models such as Precision, Recall, F1-score, and the Area under Curves of Receiver Operating Characteristic (AUC@ROC) (Xu et al., 2017). Moreover, we also employed the Top-K metrics that were proposed in Lakkaraju et al. (2015) to evaluate the model's performance in identifying the Top-ranked at-risk students based on the predicted dropout probabilities. This evaluation approach is particularly useful in situations when instructors need to allocate the limited learning resources to the students who are most in need (Hlosta et al., 2017).

Interpretable Results. While building models that can identify at-risk students accurately is a critical step in the practice of increasing students' retention rates, it is equally important to provide interpretable results that can support instructors design actionable interventions to facilitate students' learning. In the interest of converting the predicted output to actionable insights, this paper uses the Locally Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016), which is a model agnostic approach that trains an interpretable linear alternative of the original model to provide locally faithful explanations. In this paper, we use LIME to interpret the results produced during certain identified critical weeks to explain how instructors can leverage the insights drawn from the output in their decision-making processes.

Table 3. Statistical Summarizes of the Course Data.

Course type	# of students	# of dropout	# of resources	Sum of clickstreams	Observations span T(day)	Snapshot window (day)	Target period (week)
Social Science 2013J	383	60	9	648,494	268	7	5–35
Social Science 2014J	365	66	9	598,158	269	7	5–35
Stem 2014B	2498	1049	9	1,889,177	241	7	5–35
Stem 2014J	1936	898	9	2,792,972	269	7	5–35

Experimental Results

We perform the task of identifying at-risk students on the four courses in relation to the two subjects: social science and stem, respectively. Table 3 summarizes the statistical information of the dataset. The objective of the proposed model is to predict whether a student will drop out of the courses based on the demographic information and learning behaviors, as well as calculate the associated dropout probability for each student during each week. The model is first trained on the dataset from the first four weeks and then used to predict students' dropout rate beginning from the 5th week till the 35th week. During each week, there is a huge imbalance in the number of students who choose to drop out of the course and stay in the course, and this can result in a bad performance of the machine learning models. As such, we employed the Synthetic Minority Over-sampling (SMOTE) technique (Chawla et al., 2002) to rebalance the dataset during each week. And finally, the models are trained using the balanced dataset. Specifically, we set 80% of the samples in the dataset during each week as the training dataset and the remaining as the testing dataset. In the training dataset, we use the 8-fold cross-validation approach to obtain the optimal set of hyperparameters for the best accuracy of the models. After that, the final performances of the respective trained models are compared and evaluated on the testing dataset.

Comparing the Models on AUCs Weekly. The AUC metric is commonly used to assess machine learning models' performance on imbalanced datasets (Wardhani et al., 2019). In this section, we compared the AUCs of the baseline models and that of the proposed model at each week of the courses from both the social science subject and stem fields. Each subfigure compares a model's AUCs on identifying the at-risk students from the two courses of a specific subject.

Figure 5 shows the differences in the AUCs of each model in identifying at-risk students from the two courses in the social science field. Each subfigure compares a model's AUCs on the two courses each week. Comparing the

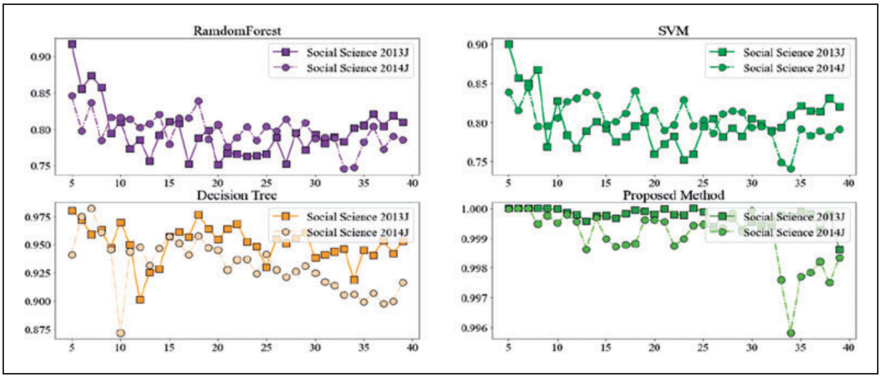

Figure 5. The Weekly AUCs of Models on Identifying At-Risk Students in the Social Science Courses.

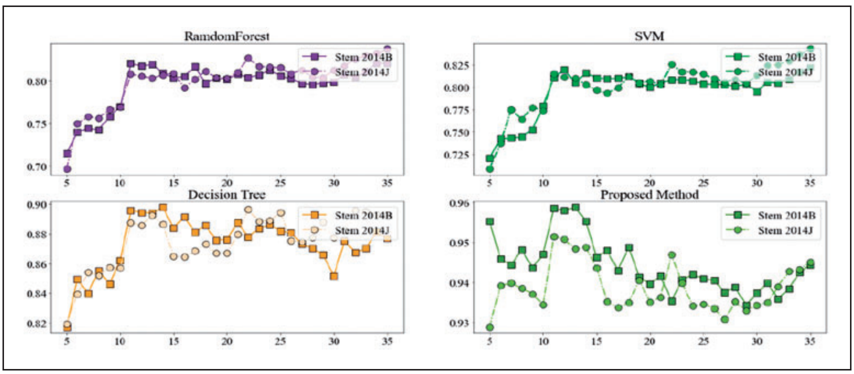

Figure 6. The Weekly AUCs of Models on Identifying At-Risk Students in the Stem Courses.

differences of AUCs on each week in each subfigure, it can be found that the maximum AUC difference is less than 0.100, which means that all the models have a stable performance on the courses from the social science field. Looking more closely at the performance of the proposed model, the maximum difference is about 0.04, which indicates that the proposed model has a more stable performance when working on similar courses.

Similar to Figure 5, Figure 6 compares the performances of the models on the two courses from the stem field. Obvious differences are found in the trends of the weekly AUCs of the baseline models comparing with those shown in Figure 5. While as for the proposed model, no significant differences in the performances are found. This further indicates that the proposed model can maintain a stable performance even working on unstable datasets. Just as the dataset about the stem courses in this study, the relatively large number of students drop out of the courses each week (see Figure 4) can cause huge differences in the distributions of samples from week to week. Moreover, examining the differences in the models' performances on the consecutive weeks, it can also be found that the variations of AUCs of the proposed model are also smaller than that of the baseline models.

The Average Performance Over the Courses. To further evaluate the performance of the proposed model in identifying the at-risk students, in this section, we compared the models based on the metrics such as precision, recall, accuracy, and f1-score. And the final performances are calculated by averaging the weekly performances measured via each of the metrics. Table 4 presents the averaged performances of each model on the dataset, with the bolded values indicating our model's performance under each metric. The results indicated an obvious superiority of the proposed model in terms of the overall performance (i.e., precision, recall, accuracy, and f1-score) in identifying at-risk students in both of the subjects.

The Weekly Top-K Performance of the Proposed Model. In many cases, instructors do not have to provide interventions for every identified at-risk student, especially in situations where there are limited supportive learning resources available. In this section, we adopted the Top-K metrics to evaluate the proposed model's

Table 4 .The Average Performance of the Models Over the Course.

Category	Method	Precision	Recall	Accuracy	F1Score
Social Science 2013J	Random forest	0.6956	0.8087	0.7275	0.7478
	SVM	0.8746	0.9324	0.8977	0.9009
	Decision tree	0.6960	0.8595	0.7423	0.7689
	Proposed approach	**0.9030**	**0.9634**	**0.9132**	**0.9322**
Social Science 2014J	Random forest	0.7060	0.8331	0.7432	0.7640
	SVM	0.8505	0.8810	0.8609	0.8637
	Decision tree	0.7035	0.8948	0.7588	0.7874
	Proposed approach	**0.9262**	**0.9000**	**0.9380**	**0.9129**
Stem 2014B	Random forest	0.7061	0.8316	0.7429	0.7636
	SVM	0.7741	0.8584	0.8023	0.8126
	Decision tree	0.6969	0.8526	0.7409	0.7667
	Proposed approach	**0.8359**	**0.9092**	**0.8649**	**0.8707**
Stem 2014J	Random forest	0.7177	0.8309	0.7519	0.7701
	SVM	0.7856	0.8342	0.8024	0.8085
	Decision tree	0.7110	0.8480	0.7514	0.7734
	Proposed approach	**0.8496**	**0.8741**	**0.8596**	**0.8615**

Bold values indicates our model's performance under each metric.

performance in identifying the top *K* at-risk students each week. Specifically, we examined the model's performance in terms of recalls and precisions under different *K*s across the course. For example, to calculate the Top-*K* recall at a specific week, we first rank the students according to the predicted dropout probabilities; Then, the students will be relabeled as 1 or 0 (i.e., dropout or not) based on the threshold determined in the problem definition section; Finally, we calculate the recall from the Top-*K* percent identified at-risk students. The calculation of Top-*K* precision follows the same steps except for the last step that calculating the precision from the Top-*K* percent identified at-risk students.

Figure 7 shows the model's performance on Top-*K* metrics when working on the social science courses with $K = 1,6,11,16$ respectively. According to Figure 4 and Table 3, we can find that the number of students who dropped out of the social science courses is relatively small during each week. As such, we set relatively small *K*s to examine the model's performance based on the Top-*K* metrics. From the figure, it can be seen that the model exhibits poor performance on both the Top-*1* recall and precision over the courses in the social science field. This indicates that the number of the Top-*1* percent of predicted at-risk students is smaller than the actual number of dropout students each week.

Looking at the model's performance on the other metrics, it can be found that the model's performances on the Top-*6* and Top-*11* metrics begin to drop after week 15 and week 28, respectively. As such, it can be concluded that we can confidently allocate the limited learning resources to at least the Top-*6* percent of the identified at-risk students before week 15, and Top-*11* percent of the at-risk students before week 28.

According to Figure 4 and Table 3, in the courses from the stem field, there were a large number of students who dropped out each week. As such, we set larger *K*s (i.e., 10, 20, 30, 40, and 50) to evaluate the model's week performance on the Top-*K* metrics. Similarly, it can be found that the model has a consistently low performance on the Top-*10* metrics across the courses. Moreover, the

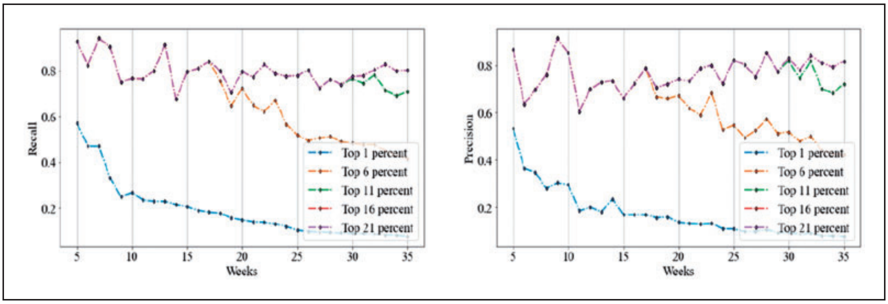

Figure 7. The Model's Weekly Performance in Terms of the Top-*K* Metrics in Identifying At-Risk Students in the Courses From the Social Science Field.

model has the same Top-*K* performance before week 9 with a relatively low recall value of around 0.55. This means that the model can only identify 55% of the actual at-risk students at week 9 even when the *K* is set as 50, which also indicates that there are still a large number of students who will drop out after this week. While at Week 19, the recall values rise to above 0.70 under the condition that *K* is larger than 30. As such, in the stem courses, the instructors can confidently allocate the limited learning resources to the Top-*30* percent identified at-risk students at Week 19.

Actionable Insights From the Interpretable Model. In this section, we attempt to interpret the model's output using the LIME approach. As described above, instructors are recommended to provide interventions at week 15 and week 25 in the social science courses and at week 19 in the stem courses according to the percentage of the at-risk students. More specifically, we select 3 representative at-risk students identified by the model in each of these weeks and interpret the model's predictive output using LIME. 8 features that most influence the model's predictive result for each student are examined. The features in orange support class 1 (i.e., dropout) while those in blue support class 0 (i.e., stay in). And they are listed in descending order according to their respective importance to the predictive model.

Drawing on the predicted dropout probabilities of the 3 students and the distributions of the critical features in each prediction in Figure 9A, it can be found that spending much time on forum discussions (i.e., forum, ou_Collaborate) and the content unrelated resources (i.e., homepage, subpage, URL) might be the major factors that lead the students to drop out of the courses. These learning behaviors also imply that these students might have certain difficulties in locating the correct learning resources for solving the problems at hand. As such, providing students with direct instructions about the necessary resources of the problem is a potentially effective intervention strategy. Similarly, in Figure 9B, students' activities related to the course content (i.e., glossary, ouContent, activityCount) are encouraged to improve the retention rate.

Figure 10 presents the distributions of the critical features that the model used for predicting 3 at-risk students with a dropout probability of 0.91, 0.88, and 0.85, respectively in week 19 of the courses from the stem field. In contrast, from the figure, it can be found that the features in terms of forum discussions (i.e., ou_Collaborate) and course content-related resources (i.e., page, resources) are more likely to support students to stay in the courses. This also means that students in this week were engaged in learning activities that require frequent collaborations with others. As such, the strategies that can facilitate student collaboration processes are favored to support students' learning in this situation.

Discussion

Our paper proposed an algorithm for identifying the students who are at risk to drop out of a course in online learning settings, with a particular focus on the accountability and interpretability of the predictive output. It identified some important insights, strategies, and practices that can be used by researchers and educators in improving the retention rates and providing students with informed interventions. Specifically, in this study, we converted the model's output from the previous week into meaningful references that can be used to infer the predictive results in the current week. Through comparisons with the performance of the baseline models, we found that considering the underlying learning associations across weeks can produce the predictive curves that are more in line with students' learning processes and further indicate good performance in identifying at-risk students. And this also converges with the findings in the study of He et al. (2015) that the smoothed curves for dropout probabilities can greatly mitigate the influences from the differences in students' learning behaviors. Moreover, the comparable performance of the model on the two different subjects also indicated the generalizability of the approach in capturing students' learning patterns in different contexts. Finally, this study goes beyond the simple binary classification by employing interpretation approaches such as LIME to further demonstrate the feasibility of building a fully automatic pipeline from identifying at-risk students to producing actionable insights in supporting their learning processes.

An imbalanced dataset can negatively impact the performance of the machine learning models because the patterns exhibited in the majority class in the dataset are more likely to dominate the overall trend, which poses great challenges for machine learning models to underpin the patterns in minority classes (Hlosta et al., 2017). Generally, the number of students likely to drop out of a course at any stage should be in a minority group, in this case, the involved datasets are more likely to be imbalanced. As such, we employed the SMOTE technique on student learning data each week to make sure samples in the two target classes in the dataset are equal. This is in line with the research of Goel and Goyal (2020). Multiple approaches (i.e., Undersampling, Oversampling, and Hybrid methods) were employed by the authors to balance the different classes in the dataset to ensure that the model could better capture the underlying patterns. Moreover, the data rebalance approaches can also help to bring to the fore the information that is hidden in the underrepresented samples in the dataset.

Related to optimally allocating the limited learning resources to the students who are most at-risk to drop out of a course, we further employed the Top-K metrics to explicitly examine the recall and precision of the model in terms of identifying the top K percent students who are most at-risk to drop out. Particularly, our study also partially answered the question brought up by

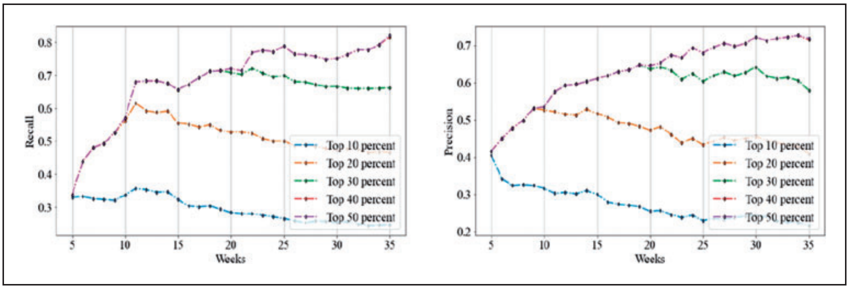

Figure 8. The Model's Weekly Performance on the Top-*K* Metrics in Identifying At-Risk Students in the Courses From the Stem Subject.

Chen et al. (2019) about whether instructors should provide interventions for all at-risk students. By examining the recalls and precisions under different *K*s each week, the optimal *K*s and weeks for allocating remedy resources or providing urgent interventions can be easily determined. According to Figure 7 and Figure 8, the best time for providing interventions, on the one hand, should be as early as possible so that the students could have more time to adjust their learning strategies to succeed in the course. On the other hand, the corresponding recall and precision should be larger than a specific threshold to ensure the credibility of the predicted results. Although the *K*s in this paper seem to be chosen arbitrarily, it is related to the number of available learning resources that can be allocated to students in real settings. As such, it is possible to first determine the *K*s based on the available critical resources in each week and then select the optimal week for providing interventions through assessing the recalls and precisions in identifying the top *K* percent students who are identified to be at risk to drop out by the model.

Our study also employed interpretation approaches to convert the predicted results into actionable insights in support of making informed decisions about intervention strategies. A particular model agnostic approach named LIME was adopted to visualize the role of the top 8 important features for generating the specific predictions. Figure 9 and Figure 10 provided the instance level interpretations about how the different combinations of the features lead students to drop out of the courses. Particularly, by examining the feature importance, instructors can gain insights about students' learning status in the corresponding week. For example, in Figure 9A, it can be found that the learning activities (i.e., glossary, dataplus) directly related to the course content can support students to stay in the course, while the activities such as viewing forums and online forums could be more likely to cause students to drop out of the course. Taken together, the learning patterns of these students indicate that they might have difficulties

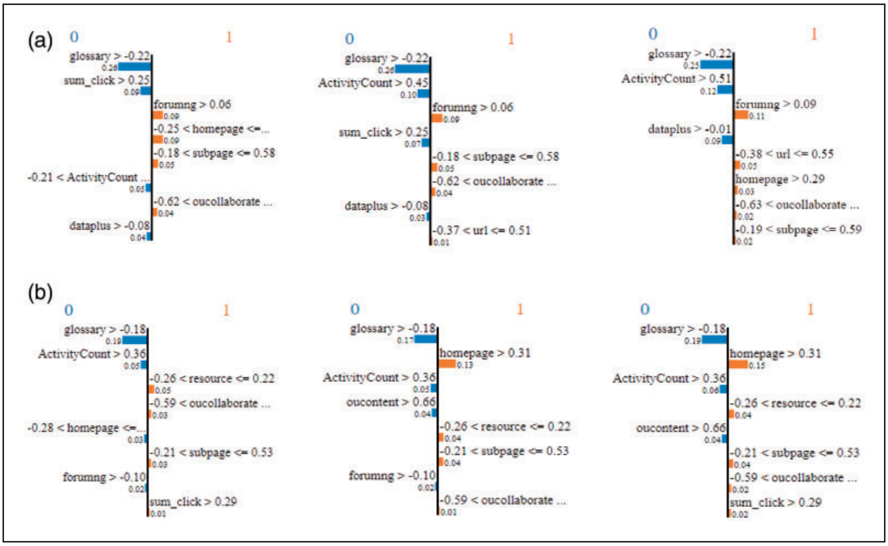

Figure 9. The LIME Interpretations About the Representative At-Risk Students in the Social Science Subject. A: Three at-risk students identified in Week 15 with predicted dropout probabilities of 0.96, 0.86, and 0.79, respectively. B: Three at-risk students identified in Week 28 with predicted dropout probabilities of 0.96, 0.77, and 0.66, respectively.

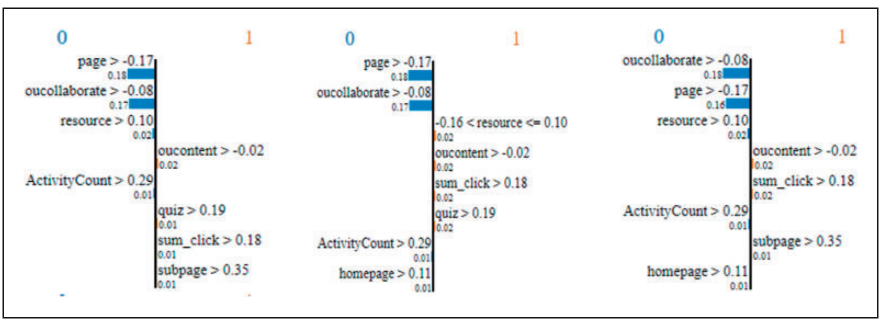

Figure 10. LIME Interpretations About the Predictions of the Representative Students in the Stem Subject.

in finding the correct resources by themselves, so they spent more time in forums seeking related help. Comparing Figure 9 and Figure 10, different learning patterns are found to correlate with students' learning states. Figure 9 indicates the content-related activities could help students to stay in the courses, while in Figure 10 it can be found that collaborative abilities can significantly contribute to students' stay in the course. As such, by combining the information drawn on the predicted results and the actual contexts of the courses, instructors can be able to provide more actionable interventions to support students' individualized learning.

Conclusion

At-risk student prediction is an essential prerequisite for providing personalized interventions to improve the retention rates in online learning settings. In this paper, we proposed a novel model that can convert the predicted dropout probabilities in the previous week into meaningful references for producing accountable predictions in the current week. The algorithm was conducted on two selected subjects (i.e., social science and stem) from the OULAD dataset. The effectiveness of the proposed algorithm was evaluated from the following aspects: (1) comparing the weekly AUCs when identifying at-risk students on different courses; (2) comparing the averaged performance over the courses in terms of precision, recall, accuracy, and f1-score; (3) evaluating the proposed model using the Top-*K* metrics; (4) using LIME to interpret the predicted results to provide instructors with actionable insights. The experiment indicates the proposed model can deliver the best performance compared with the baseline models when identifying at-risk students in terms of both weekly AUCs and the averaged performance over the courses. Moreover, the combination of Top-*K* performance evaluations and the LIME approaches further demonstrated the feasibility of proposing an accountable pipeline for automatically providing individualized interventions for at-risk students in online learning settings.

While our work presents a complete and actionable pipeline in terms of providing personalized interventions for at-risk students, a potential aspect that should be noted is the "fairness of the interventions". This requires that the interventions should not be impacted by the differences in students' demographic information. For example, as shown in Figure 2, there is a significant difference in the number of students with an "A Level or Equivalent" degree and those with a "No Formal Equals" degree in both of the subjects. And the number of male students in the stem subject is significantly larger than that of the female students. These imbalances in demographic information could introduce "unintentional discriminations" in the models used for identifying at-risk students and producing interpretable results, with the produced output cannot represent the actual patterns that existed in unprivileged groups (Kizilcec & Lee, 2020). As such, a potential direction of future research is to examine the fairness of algorithmic recommendations generated for students with different demographic backgrounds (i.e., genders, ages, races, educational levels). Particularly, questions such as "*Whether students with similar learning statues are received similar treatments?*", "*To which extent the treatments are independent of students' demographic information?*", and "*How to ensure the similar* *improvement in students' learning performance with similar treatments imposed on?*" should be further investigated in future studies.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Bo Pei https://orcid.org/0000-0002-6328-6929
Wanli Xing https://orcid.org/0000-0002-1446-889X

References

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access*, *9*, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

Aljohani, N. R., Fayoumi, A., & Hassan, S.-U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability*, *11*(24), 7238. https://doi.org/10.3390/su11247238

Carvalho, P. F., Sana, F., & Yan, V. X. (2020). Self-regulated spacing in a massive open online course is related to better learning. *NPJ Science of Learning*, *5*(1), 2. https://doi.org/10.1038/s41539-020-0061-1

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. *Mathematical Problems in Engineering*, *2019*, 1–11. https://doi.org/10.1155/2019/8404653

Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, *135*, 113325. https://doi.org/10.1016/j.dss.2020.113325

Fei, M., & Yeung, D.-Y. (2015). Temporal models for predicting student dropout in massive open online courses. In P. Cui, J. Dy, C. Aggarwal, Z. Zhou, A. Tuzhilin, H. Xiong, & X. Wu (Eds.), *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 256–263). IEEE. https://doi.org/10.1109/ICDMW.2015.174

Goel, Y., & Goyal, R. (2020). On the effectiveness of self-training in MOOC dropout prediction. *Open Computer Science*, *10*(1), 246–258. https://doi.org/10.1515/comp-2020-0153

Gunnarsson, B. L., & Alterman, R. (2012). Predicting failure: A case study in co-blogging. In *Proceedings of the 2nd international conference on learning analytics and knowledge—LAK' ' '12* (p. 263). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330665

Hassan, S., Waheed, H., Aljohani, N. R., Ali, M., Ventura, S., & Herrera, F. (2019). Virtual learning environment to predict withdrawal by leveraging deep learning. *International Journal of Intelligent Systems*, *34*(8), 1935–1952. https://doi.org/10.1002/int.22129

He, J., Baileyt, J., Rubinstein, B. I. P., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. *Proceedings of the National Conference on Artificial Intelligence*, *3*, 1749–1755.

Hidalgo, Á. C., Ger, P. M., & Valentín, L. D. L. F. (2021). Using meta-learning to predict student performance in virtual learning environments. *Applied Intelligence*. Advance online publication. https://doi.org/10.1007/s10489-021-02613-x

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *Proceedings of the seventh international learning analytics & knowledge conference on - LAK' ' '* (Vol. 17, pp. 6–15). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027449

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2018). Are we meeting a deadline? Classification goal achievement in time in the presence of imbalanced data. *Knowledge-Based Systems*, *160*, 278–295. https://doi.org/10.1016/j.knosys.2018.07.021

Hooshyar, D., Pedaste, M., & Yang, Y. (2019). Mining educational data to predict students' performance through procrastination behavior. *Entropy (Entropy)*, *22*(1), 12. https://doi.org/10.3390/e22010012

Jayaprakash, S., Krishnan, S., & Jaiganesh, V. (2020). Predicting students academic performance using an improved random forest classifier. In *2020 international conference on emerging smart computing and informatics (ESCI)* (pp. 238–243). https://doi.org/10.1109/ESCI48226.2020.9167547

Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. *arXiv preprint* arXiv:2007.05443.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. *Scientific Data*, *4*, 170171. https://doi.org/10.1038/sdata.2017.171

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining—KDD' ' '* (Vol. 15, pp. 1909–1918). ACM. https://doi.org/10.1145/2783258.2788620

Leite, W. L., Jing, Z., Kuang, H., Kim, D., & Huggins-Manley, A. C. (2021). Multilevel mixture modeling with propensity score weights for quasi-experimental evaluation of virtual learning environments. *Structural Equation Modeling: A Multidisciplinary Journal*. Advance online publication. https://doi.org/10.1080/10705511.2021.1919895

Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016). Big data application in education: Dropout prediction in EDX MOOCS. In *2016 IEEE second international conference on multimedia big data (BigMM)* (pp. 440–443). IEEE. https://doi.org/10.1109/BigMM.2016.70

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

Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2020.1727529

Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear support vector machines for prediction of student performance in school-based education. *Mathematical Problems in Engineering*, *2020*, 1–7. https://doi.org/10.1155/2020/4761468

Onah, D. F. O., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses : Behavioural patterns MOOC dropout and completion: Existing evaluations. In L. G. Chova, A. L. Martinez, & I. Torres (Eds.), *Proceedings of the 6th international conference on education and new learning technologies (EDULEARN14)* (pp. 1–10). IATED Academy.

Qiao, C., & Hu, X. (2020). A joint neural network model for combining heterogeneous user data sources: An example of at-risk student prediction. *Journal of the Association for Information Science and Technology*, *71*(10), 1192–1204. https://doi.org/10.1002/asi.24322

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

Rovira, S., Puertas, E., & Igual, L. (2017). Data-driven system to predict academic grades and dropout. *PLoS One*, *12*(2), e0171207. https://doi.org/10.1371/journal.pone.0171207

Tan, F., Wei, Z., He, J., Wu, X., Peng, B., Liu, H., & Yan, Z. (2018). A blended deep learning approach for predicting user intended actions. In *2018 IEEE international conference on data mining (ICDM)* (pp. 487–496). IEEE. https://doi.org/10.1109/ICDM.2018.00064

Tang, H., Xing, W., & Pei, B. (2019). Time really matters: Understanding the temporal dimension of online learning using educational data mining. *Journal of Educational Computing Research, 57*(5), 1326–1347.

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

Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., & Lestantyo, P. (2019). Cross-validation metrics for evaluating classification performance on imbalanced data. In *2019 international conference on computer, control, informatics and its applications (IC3INA)* (pp. 14–18). https://doi.org/10.1109/IC3INA48034.2019.8949568

Xing, W., Pei, B., Li, S., Chen, G., & Xie, C. (2019). Using learning analytics to support students' engineering design: the angle of prediction. *Interactive Learning Environments, 2019*, 1–18.

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research, 57*(3), 547–570.

Xing, W., Tang, H., & Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. *The Internet and Higher Education, 43*, 100690.

Xiong, F., Zou, K., Liu, Z., & Wang, H. (2019). Predicting learning status in MOOCs using LSTM. In *Proceedings of the ACM turing celebration conference—China on' '—ACM TURC '19* (pp. 1–5). ACM. https://doi.org/10.1145/3321408.3322855

Xu, J., Moon, K. H., & van der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), 742–753. https://doi.org/10.1109/JSTSP.2017.2692560

Zhang, J.-H., Zou, L., Miao, J., Zhang, Y.-X., Hwang, G.-J., & Zhu, Y. (2020). An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, *28*(2), 231–215. https://doi.org/10.1080/10494820.2019.1636078

Author Biographies

Bo Pei is a PhD candidate in School of Teaching & Learning, College of Education at the University of Florida. His research interests are STEM education, educational data visualization, and machine learning.

Wanli Xing is an Assistant Professor of Educational Technology at the University of Florida. His research interests are artificial intelligence, learning analytics, STEM education, and online learning.