applied sciences

MDPI

Article

Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach

Zhaoyu Shou [1,*], Mingquan Xie [2], Jianwen Mo [2] and Huibing Zhang [3]

[1] Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory, Guilin University of Electronic Technology, Guilin 541004, China
[2] School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; 21022303168@mails.guet.edu.cn (M.X.); jwmo@guet.edu.cn (J.M.)
[3] School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; zhanghuibing@guet.edu.cn
* Correspondence: guilinshou@guet.edu.cn

Abstract: As an emerging teaching method, online learning is becoming increasingly popular among learners. However, one of the major drawbacks of this learning style is the lack of effective communication and feedback, which can lead to a higher risk of students failing or dropping out. In response to this challenge, this paper proposes a student performance prediction model based on multidimensional time-series data analysis by considering multidimensional data such as students' learning behaviors, assessment scores, and demographic information, which is able to extract the characteristics of students' learning behaviors and capture the connection between multiple characteristics to better explore the impact of multiple factors on students' performance. The model proposed in this paper helps teachers to individualize education for students at different levels of proficiency and identifies at-risk students as early as possible to help teachers intervene in a timely manner. In experiments on the Open University Learning Analytics Dataset (OULAD), the model achieved 74% accuracy and 73% F1 scores in a four-category prediction task and was able to achieve 99.08% accuracy and 99.08% F1 scores in an early risk prediction task. Compared with the benchmark model, both the multi-classification prediction ability and the early prediction ability, the model in this paper has a better performance.

Keywords: online learning; multidimensional time-series data; student performance prediction; multi-classification prediction; individualized education; early prediction

check for updates

Citation: Shou, Z.; Xie, M.; Mo, J.; Zhang, H. Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach. *Appl. Sci.* **2024**, *14*, 2522. https://doi.org/10.3390/app14062522

Academic Editor: Luis Javier Garcia Villalba

Received: 19 February 2024
Revised: 14 March 2024
Accepted: 15 March 2024
Published: 16 March 2024

1. Introduction

With the continuous improvement of online learning platforms, a proliferation of schools has opened up online course teaching, providing students with more flexible learning methods. However, there is a significant problem with the online learning process: due to the lack of face-to-face interaction between instructors and students, it is difficult for instructors to accurately grasp the level of understanding and mastery of course content, which may lead to students' poor performance on exams. Therefore, it is particularly important to capture the hidden information related to student performance from massive educational data collected by online learning platforms. Using this information, students at risk of failing and dropping out of school can be identified and an intervention can be carried out as early as possible to effectively improve their learning status and enhance learning effectiveness [1]. Data mining techniques aid researchers in discovering and understanding latent critical information within massive educational datasets, enabling the identification and prediction of students' performance trends [2]. Hughes et al. [3] mined high/active engagement, attendance, and multiple interactions as key factors for students to achieve higher marks on a Massive Open Online Learning

(MOOC) platform by using data mining techniques. Such information extraction is not confined solely to the field of data mining but also extends into the domain of learning analytics. With the rise in learning analytics, there has been a growing emphasis on investigating student behaviors and aggregating various methodologies to enhance the comprehension of student behaviors, consequently enabling the prediction of students' academic performance [4]. As a category of data reflecting student behaviors, clickstream data illustrate the pathways students take while navigating one or multiple learning websites on an online learning platform [5]. Casalino et al. [6] used data from students' interactions with the virtual learning environment to analyze learning through a neuro-fuzzy system, showing that students who regularly visit the course homepage and actively participate in the teaching and learning activities through evaluation tests are more likely to pass the exams. In addition to clickstream data, students' demographic information and assessment scores have also been verified to be closely correlated with students' academic performance [7,8]. Considering more dimensions of information beyond students' learning behaviors, modeling and analyzing students' academic performance as a whole can more accurately predict students' academic performance.

The early prediction of student performance allows for the proactive identification of at-risk students, enabling instructors to devise additional courses, and assignments, or implement other measures to enhance students' performance before the conclusion of the course. Currently, most research only applies to predicting student performance at the end of the course, which does not allow for timely intervention. Therefore, it is particularly important to predict student performance as early as possible before the end of the course [9]. Furthermore, most studies have predicted student performance (pass, fail) on a dichotomous basis, which would result in students at an excellent level losing their upward mobility and failing to maximize benefits from the course [10]. To promote the sustainable development of online education and provide a basis for teachers to intervene with at-risk students promptly, this paper synthesizes multidimensional data, such as students' learning behaviors, assessment scores, and demographic information, and proposes a prediction model for student performance based on the analysis of multidimensional time-series data (MTAPSP), aiming at efficiently and accurately predicting students' performances in an online learning platform. The main contributions of this study can be summarized in the following three points:

1. A student performance prediction model for multi-dimensional time-series data analysis is proposed, which uses the multi-head self-attention (MHSA) mechanism to better integrate features such as time-series behaviors, assessment scores, and demographic information, avoiding the limitations of single-dimensional analysis while enhancing the model's ability to predict students' performance in multi-classification.
2. In this paper, we use temporal behavioral features extracted from students' chronological behavior and assessment scores using multilayer LSTM and further enhance the model's nonlinear mapping ability to these features through ANN, which improves the model's early prediction ability to identify at-risk students as early as possible and assists teachers in making timely interventions.
3. A large number of experiments have been conducted on the publicly available dataset OULAD, and the experimental results demonstrate that the model proposed in this paper outperforms the benchmark model in terms of both multi-classification prediction and early prediction ability.

The paper is organized as follows: Section 2 provides an overview of related research work. Section 3 presents the dataset used for experiments. Section 4 provides a detailed description of the multi-dimensional time-series analysis-based student performance prediction model (MTAPSP). Section 5 describes the experimental process and analyzes the results to evaluate the performance of the proposed algorithm. Finally, Section 6 summarizes the paper and proposes future research.

2. Related Work

Student performance prediction research can be categorized into regression and classification problems, where the former predicts a score value, and the latter predicts a category [11]. Earlier studies in student performance prediction focused on traditional machine learning methods, which used logistic regression [12], support vector machines [13], and decision trees [14] to establish models for predicting student performance. In a study [15], an automatic method for observing and predicting student grades was proposed. This method utilized a genetic algorithm to capture the 30 best attributes from students' historical learning data and trained a K-NN regression model and a decision tree using these features and labels to predict students' performance score and categories. PEK et al. [16] used naive Bayes, random forest, decision tree, AdaBoost classifier, logistic regression, and KNN algorithms as basic learners, support vector machine (SVM) as meta-learner, and created a stacking method to develop a hybrid ensemble model. By analyzing the data, they discovered important features that influence student learning outcomes and successfully helped teachers identify students at risk. Jawad et al. [17] and Bujang et al. [18] combined the SMOTE technique with machine learning techniques to improve the impact of data imbalance on the model and enhance the accuracy of student performance prediction. Hung et al. [19] proposed a method for predicting student performance based on time-series clustering. The method aggregates learning behavior data such as the frequency of accessing course materials, frequency of reading forums, number of discussions, and number of replies posted to identify at-risk students and predicts student performance more accurately than traditional frequency aggregation methods. Traditional machine learning methods ignore time information in the original data, which cannot effectively capture the impact of time features on student performance. However, deep neural networks can effectively address the problem of time information missing [20]. Therefore, some researchers have begun to utilize deep learning techniques to predict student performance. For example, He et al. [21] used a GRU network to extract time-series features from clickstream data and assessment score data and combined them with demographic features to predict student performance. Liu et al. [22] proposed a hybrid deep learning model that can extract time-behavioral and overall behavioral information from learning behavior data to more accurately predict high-risk students. Qu et al. [23] constructed a student performance prediction framework with an attention mechanism, in which an LSTM neural network was used to reflect students' learning processes, and a DSP-based adapter was used to enhance the importance of key information and improve the accuracy of student performance prediction. Kusumawardani et al. [24] proposed a transformer-based method for predicting student performance by converting the learning behavior data of students into a sequential feature vector. In some studies [25–27], authors used convolutional neural networks (CNN) to extract high-dimensional time information from the time series of student activities to better extract spatio-temporal features and utilized LSTM to capture the sequence information of student learning dynamics to more accurately predict student performance. Chen et al. [28] proposed a performance prediction model based on the differences in patterns of student behavioral features for students with significant performance changes, which uses a multi-head attention mechanism to automatically select more important higher-order behavioral combinations of features, maintain higher temporal accuracy, and can predict student performance more accurately. Li et al. [29] considered the potential relationships between students and used multiple graphs with different topologies to reflect the relationships between students and proposed a student performance prediction model based on a multi-topological graph neural network (MTGNN). Yang et al. [30] utilized clickstream data and assessment scores as input data to train a time series neural network to capture the unique features of each student's learning pattern. The results showed that the correlation between behavior features and scores was not very high, and it needed to consider the overall effect and nonlinear predictive factors.

In summary, most existing performance prediction models have not taken into account age group, residence location, and other demographic information, and the influence of assessment scores on student performance has not been fully considered in many of them. Additionally, many of these models have limited early prediction abilities. Most existing research has primarily focused on binary classification of student performance, with insufficient granularity in terms of dividing student performance. This paper considers additional information beyond a student's learning behavior to achieve the early prediction of student performance and multi-classification prediction of performance, providing effective support for individualized teaching.

3. Dataset

3.1. Introduction to the Dataset

The validation of the MTAPSP student performance prediction model using the Open University Learning Analytics Dataset (OULAD) provided by the Open University UK. OULAD comprises detailed learning behavior log data of 32,593 students in seven courses during the 2013–2014 academic year, including demographic data and clickstream data of student interactions in the VLE, which provided effective data support for analyzing the relationship between student behavior and student performance [31]. In the OULAD dataset, the seven courses are subdivided into 22 instructional modules covering a wide range of areas in science, social sciences, technology, engineering, and mathematics. Each module is named according to the year and month in which it is offered, e.g., a module offered in February 2013 is named 2013B, and a module offered in October 2013 is named 2013J, reflecting the fact that each module is offered multiple times during the year. Course information is provided in Table 1.

Table 1. Course information table.

Module	Domain	Presentations	Students
AAA	Social Sciences	2	748
BBB	Social Sciences	4	7909
CCC	STEM	2	4434
DDD	STEM	4	6272
EEE	STEM	3	2934
FFF	STEM	4	7762
GGG	Social Sciences	3	2534
Total	7	22	32,593

Each course has resources in the VLE, and there are a total of 20 activities for students in the VLE, and information on their activity types is shown in Table 2. Students are evaluated on the Teacher-Marked Assessment (TMA), the Computer-Marked Assessment (CMA), and the Final Exam (Exam), with the final performance categorized into four outcomes: "Distinction", "Pass", "Fail", and "Withdrawn". In addition to this, student information and student registration information is called demographic information, and the information is shown in Table 3.

Table 2. Type of learning behavior statistics.

Type of Learning Behavior	Description
Folder	Open folder
Forumng	Clicks on the discussion forum
Oucollaborate	Clicks on the online video discussions
Oucontent	Clicks on the contents of the assignment
Ouwiki	Query with Wikipedia
Ouelluminate	Participate in simulation course seminars
Quiz	Clicks on the course quiz
Questionnaire	Participate in simulation course seminars
Dataplus	Supplementary data
Dualpane	Access double window
Homepage	Clicks on the course homepage

Table 2. *Cont.*

Type of Learning Behavior	Description
Htmlactivity	Web activity
Page	Clicks on the information related to course
Resource	Clicks on the course homepage
RepeatActivity	Repetitive activity
Glossary	Access the glossary
Url	Clicks on the links to audio/video contents
Subpage	Clicks on the other sites enabled in the course
Sharedsubpage	Shared pagination
Externalquiz	Complete extracurricular quizzes

Table 3. Introduction to student demographics.

Type of Student Demographics	Description
Code_module	Module identification code
Code_presentation	Presentation identification code
Gender	Student's gender
Region	The geographic region, where the student lived while taking the module presentation
Highest Education	The highest student education level on entry to the module presentation
Imd_band	Band of statistical data
Age_band	Band of student's age
Num_of_prev_attempts	The number of times the student has attempted this module
Studied_credits	The total number of credits for the modules the student is currently studying
Disability	Indicates whether the student has declared a disability

3.2. Data Processing

In this paper, the data were preprocessed while using the OULAD dataset for model validation. First, samples with incorrect date fields were removed; then all course lengths were standardized to 270 days. To verify the early prediction ability of the model, this paper divided the data by course time length into five segments of 20%, 40%, 60%, 80%, and 100% lengths and summarized the clickstream data and weekly average click volume for each segment as representative variables for the student's virtual learning environment (VLE) interaction. In addition, to reduce the dimensionality and complexity of the model, integer encoding was used instead of the classification data in the demographic data. Finally, the clickstream data, assessment scores, and coded demographic data were concatenated to form the training data for the model.

4. MTAPSP Model

4.1. MTAPSP Model Structure

The student performance prediction model proposed in this paper consists of two parts: a multilayer LSTM network with a multi-head self-attention mechanism and an artificial neural network (ANN). First, multi-layer LSTM was used to extract time-series learning behavior features from the learning behavior time series data. Then, a multi-head self-attention mechanism was used to evaluate the importance of time-behavior features, assessment score features, and demographic features, and extract key feature information. Finally, ANN was used to integrate temporal behavioral features, assessment score features, and demographic data features to improve the performance of the model. The MTAPSP model structure is shown in Figure 1.

4.2. Multilayer LSTM Based on a Multi-Head Self-Attention Mechanism

This module aims to extract temporal behavioral features from the time-series data of learning behaviors through a three-layer LSTM. It utilizes a multi-head self-attention mechanism to capture the importance of temporal behavioral features, assessment score features, and demographic data features—three distinct dimensions—thereby further extracting more salient features.

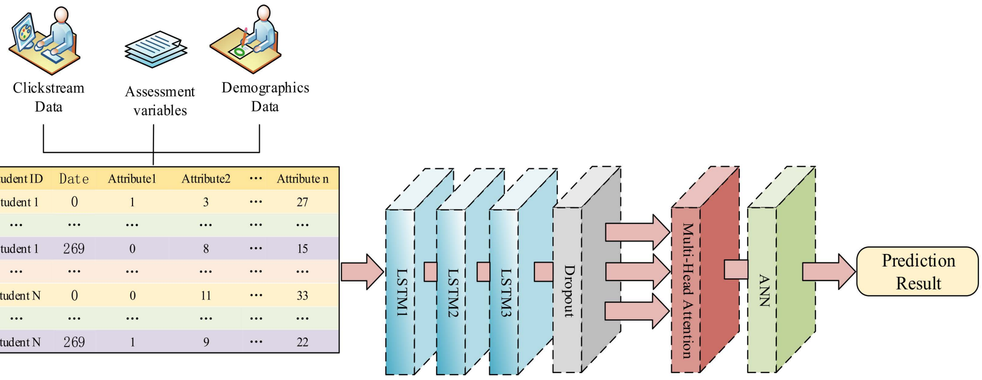


Figure 1. MTAPSP model structure.

4.2.1. Multilayer LSTM

Traditional recurrent neural networks (RNN) suffer from the problem of gradient vanishing for long sequence data like OULAD, in contrast, LSTM can efficiently convey and express the information of long-time sequences and solve the problem of long-term dependency [4]. As shown in Figure 2, compared to single-layer LSTM, a three-layer LSTM can enhance the extraction of long-distance memory and contextual information, better capture the long-distance dependencies in the sequence of students' learning time behaviors, and improve the model's generalization ability and robustness.

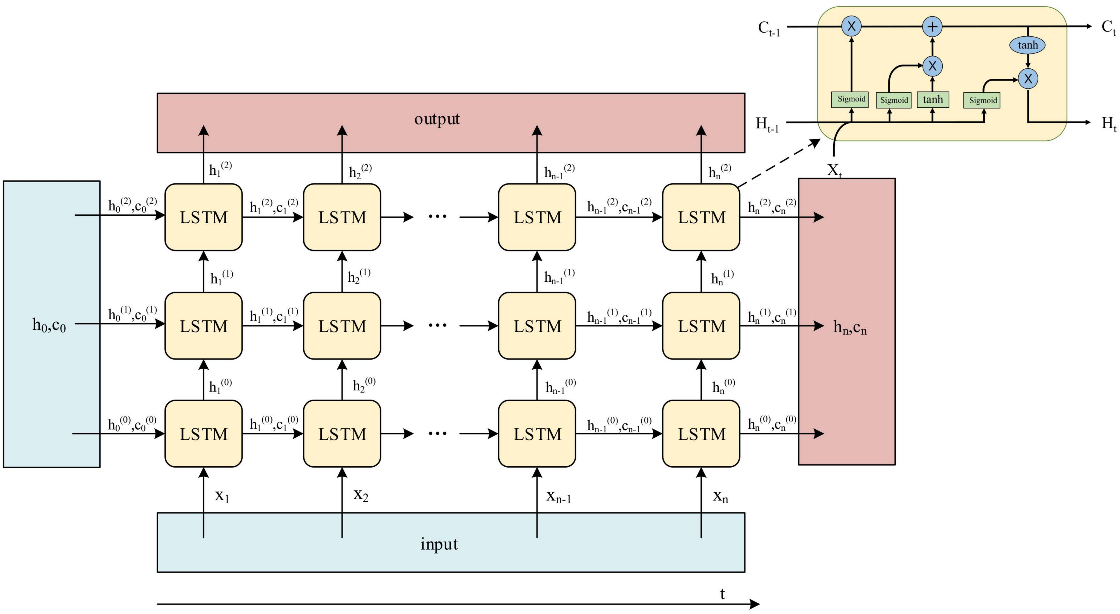


Figure 2. Three-layer LSTM network structure.

The input data $X_t \in R^{t \times m}$ is a two-dimensional tensor, where t represents the number of course days for a student, and m denotes the number of features in the input data. LSTM uses a gating mechanism to control the flow of information by feeding data X_t into the LSTM network, which sequentially goes through the computational process of forgetting gates, input gates, and output gates in LSTM. In the forget gate, LSTM integrates new input data and the output from the previous time step to determine the omission of irrelevant information. The computational formula is expressed as Equation (1).

$$f_t = Sigmoid(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

In the input gate, LSTM filters and updates the information by multiplying the current input data with the hidden state of the previous moment, which is calculated as shown in Equations (2)–(4).

$$i_t = Sigmoid(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\widetilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (4)$$

In the output gate, LSTM extracts relevant information from the vector obtained by integrating the current input data with the previous time step's output value, determining the information to be outputted. The computational formulas as shown in Equations (5) and (6).

$$o_t = Sigmoid(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

The above equations, where b_f, b_i, b_c, b_o, W_f, W_i, W_c, W_o are the biases and weights that correspond to the three gates, and denote the activation function.

4.2.2. MHSA

The self-attention mechanism enables the model to automatically learn and focus on the most relevant information within the input data [32]. A multi-head self-attention mechanism that can map a sequence into multiple specific spaces separately perform multiple attention computations, and finally stitch the results together to obtain richer contextual information.

As shown in Figure 3, in MTAPSP, when the multi-head self-attention mechanism receives the output sequence from LSTM, the sequence is linearly transformed into vector matrices Q (Query), K (Key), and V (Value), facilitating multiple attention computations. The single self-attention is calculated as shown in Equation (7).

$$\begin{aligned} Attention(Q, K, V) &= softmax(\frac{QK^T}{\sqrt{d_k}})V \\ d_k &= \frac{d_{\text{model}}}{h} \end{aligned} \quad (7)$$

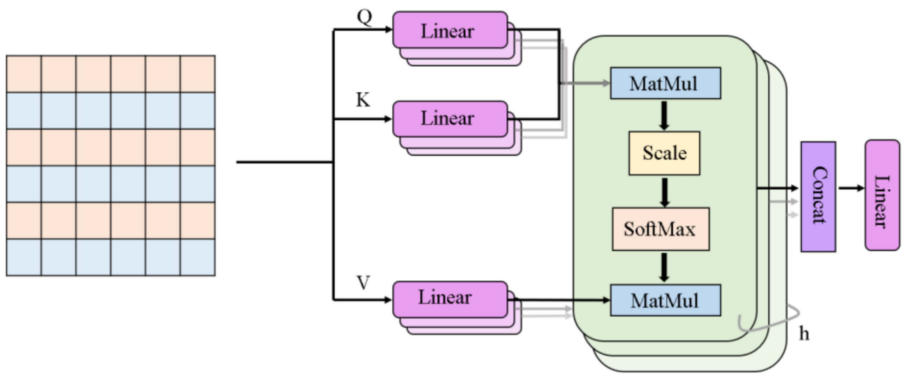


Figure 3. Structure of the multi-head self-attention mechanism.

In the equation, d_k represents the dimension of the vectors in a single attention computation, $softmax$ denotes the weight normalization function, and the product of the dimension d_k and the number of heads h in a single attention computation equals the model's dimension d_{model}. The multi-head self-attention mechanism stitches together the results of each set of single self-attention calculations to perform a linear transformation and finally output the final result, which is calculated by the following formula:

$$MultiHead(Q,K,V) = Concat(head_1,\ldots,head_h)W^o \\ head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{8}$$

where W^O, W_i^Q, W_i^K, $W_i^V \in R^{d_{model}\times d_k}$ is the weight matrix.

4.3. ANN

ANN learning is performed through multiple iterations, which prioritizes previous biases iteratively adjusting weights to classify potential outcomes [33]. Using ANN to integrate student learning behavior features, assessment score features, and demographic features, as shown in Figure 4, the algorithm can nonlinearly map the tensor resulting from computations through the multi-head attention mechanism. This process extracts correlated information between features and outputs the predicted student performance.

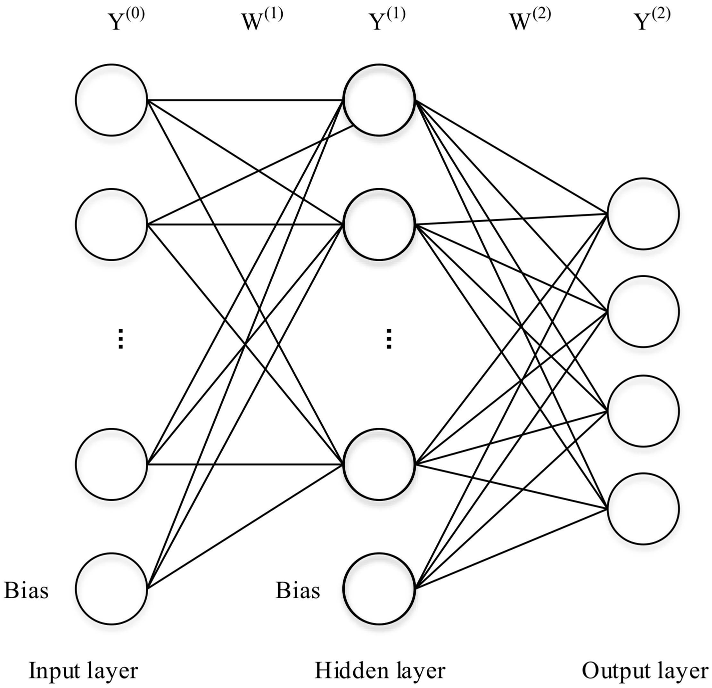


Figure 4. ANN structure.

ANN consists of an input layer, hidden layers, and an output layer. The output expression of the hidden layer is shown in Equation (9) and the output expression of the output layer is shown in Equation (10).

$$Y^{(1)} = \delta[W^{(1)}Y^{(0)} + b_1] \tag{9}$$

$$Y^{(2)} = \delta[W^{(2)}Y^{(1)} + b_2] \tag{10}$$

where δ is the activation function, $W^{(1)} \in R^{m\times r}$, $W^{(2)} \in R^{r\times n}$, m are the dimensions of the output vectors of the multi-head attention mechanism, r is the dimension of the output vectors of the hidden layer, n is the dimension of the output vectors of the output layer, and b_1, b_2 are the bias vectors.

The algorithmic flow for multi-classification prediction using the MTAPSP model is shown in Algorithm 1 below.

Algorithm 1 MTAPSP Algorithm

Input: N: Number of students; T: Number of course days; U: Student clickstream data matrix, which has a shape $N \times T \times 20$; C: Student assessment variables matrix, which has shape $N \times T \times 3$; D: Demographics data matrix, $N \times T \times 10$; K: Max epoch
Output: Y: Predicted Results of Student Performance
1: Splice U, C, D to get matrix X;
2: Initialization learning rate, hyper-parameter randomly and LSTM parameter;
3: **for** epoch $\leftarrow$ 1 to K do
4: $Y_1 \leftarrow$ LSTM1($X[n]$);
5: $Y_2 \leftarrow$ LSTM2($Y_1[n]$);
6: $Y_2 \leftarrow$ LSTM3($Y_2[n]$);
7: $Q_i, K_i, V_i \leftarrow Y_3[n]$; $//i \in [0,3]$, i is the number of heads of the multi-head self-attention mechanism
8: Atti $\leftarrow$ softmax$(\frac{Q_i K_i^T}{\sqrt{d_k}})V_i$; $//$ d_k is the dimension of X
9: $Y_{Att} \leftarrow$ concat Atti; $i \in [0,3]$;
10: $Y \leftarrow$ ANN(Y_{Att})
11: **end**
12: Return Y;

5. Experimental

5.1. Evaluation Indicators

To evaluate the performance of the MTAPSP prediction model proposed in this paper, accuracy, precision, recall, and F1-score, which are commonly used in student performance prediction models, are used in this paper as evaluation metrics [34]. The formula for accuracy is shown in Equation (11):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

where TP represents the number of correctly predicted positive classifications, FP represents the number of incorrectly predicted positive classifications, TN represents the number of correctly predicted negative classifications, and FN is the number of incorrectly predicted negative classifications. The precision rate represents the model's prediction accuracy in positive sample results, which is calculated as shown in Equation (12):

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

Recall is the probability that the model correctly predicts a positive sample among positive samples and is calculated as shown in Equation (13):

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

F1-scoreis the reconciled mean of precision and recall, which is calculated as shown in Equation (14):

$$F1\text{-}Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

5.2. Experimental Environment and Hyper-Parameter Settings

The experimental environment of this study is shown in Table 4. The dataset is randomly split into 80% training data and 20% testing data for experiments. The initial learning rate of the MTAPSP model is set to 0.001, the number of neurons in the LSTM hidden layer is set to 32, the number of heads in the multi-head attention mechanism is set to 4, the number of neurons in the hidden layer of the ANN is set to 64, the batch size of the model is set to 50, the cross-entropy loss function is used, the optimizer is set to Adam, and the number of iterations of the model is set to 100.

Table 4. Experimental environment.

Experimental Environment	Environment Configuration
Operating systems	Windows 10
CPU	Intel core i5-6700HQ
Video Cards	NVIDIA GTX 960
RAM	16 GB
DISK	1T SSD
Programming Languages	Python 3.6
Framework	Scikit-learn 0.24.2 + Pytorch 1.10.2

5.3. Experimental Results and Analysis

This section provides a detailed description of the experimental results of the MTAPSP model and the benchmark model for two scenarios: firstly, the multi-classification prediction performance. In the experiments, the "Distinction", "Fail", "Pass", and "Withdrawn" categories of samples are selected for verification. Second, the performance of the model in terms of binary prediction performance is explored, specifically validating the effectiveness of the MTAPSP model in predicting students who may be at risk in the first few weeks of school and its potential benefits. In addition, ablation experiments aimed at exploring the role and impact of MHSA and ANNs in the MTAPSP model.

5.3.1. Multi-Classification Prediction Performance

To validate the advantages of the MTAPSP model for multi-classification prediction, four benchmark models, random forest (RF), naive Bayes (NB), deep feedforward neural network (DFFNN) [34], and sequential engagement-based prediction network for academic performance (SEPN) [26], were used for comparative experiments, using the students' clickstream data, their assessment scores, and the data of the students' demographic characteristics, and each of these models' performance is shown in Table 5.

Table 5. Comparison of model multi-classification prediction performance.

Precision	**RF**	**NB**	**DFFNN**	**SEPN**	**MTAPSP**
Distinction	0.66	0.24	0.66	0.40	0.58
Fail	0.59	0.34	0.55	0.55	0.62
Pass	0.78	0.62	0.74	0.71	0.79
Withdrawn	0.73	0.72	0.76	0.74	0.78
Macro avg.	0.69	0.48	0.68	0.60	0.69
Weighted avg.	0.71	0.55	0.70	0.66	0.73
Recall	**RF**	**NB**	**DFFNN**	**SEPN**	**MTAPSP**
Distinction	0.54	0.57	0.47	0.07	0.40
Fail	0.43	0.37	0.36	0.29	0.51
Pass	0.89	0.454	0.90	0.96	0.89
Withdrawn	0.77	0.65	0.80	0.83	0.83
Macro avg.	0.66	0.51	0.63	0.54	0.66
Weighted avg.	0.72	0.50	0.71	0.70	0.74
F1-score	**RF**	**NB**	**DFFNN**	**SEPN**	**MTAPSP**
Distinction	0.59	0.34	0.55	0.12	0.47
Fail	0.50	0.36	0.44	0.38	0.56
Pass	0.83	0.52	0.81	0.82	0.84
Withdrawn	0.75	0.68	0.78	0.79	0.80
Macro avg.	0.67	0.47	0.64	0.53	0.67
Weighted avg.	0.71	0.52	0.70	0.65	0.73
Accuracy	0.72	0.50	0.71	0.70	0.74

Table 5 shows that the MTAPSP model obtained the best performance in most cases. Although the RF model performs best in differentiating between "Distinction" performance categories, given the imbalance of the data, we must consider evaluation metrics based on F1 scores. The MTAPSP model has slightly higher F1 scores than the RF model for the categories "Failed", "Passed", and "Withdrawn", indicating that the MTAPSP model has better predictive power in the face of data imbalance. Compared to the deep learning models DFFNN and SEPN, the MTAPSP model achieved superior results in all metrics, which confirms that it is more capable of capturing temporal features of learning behaviors and integrating multidimensional features to more accurately predict student performance.

The multi-classification prediction experiments clearly demonstrate the excellent performance of the MTAPSP model in this area. Teachers can predict student performance by utilizing the MTAPSP model and are able to gain insight into the academic level of their students and develop targeted instructional programs accordingly. For instance, students who are predicted to have "Distinction" results are recommended higher-level courses to deepen their learning and improve their abilities; students who are predicted to withdraw from school are provided with the necessary support and assistance by finding out whether they have any difficulties in their academic life. The application of the MTAPSP model not only provides an effective tool for predicting student performance but also provides a powerful means of guidance for educators, enabling them to counsel students in a more comprehensive and individualized manner, and promoting the improvement of teaching quality.

5.3.2. Early Risk Prediction

To validate the performance of the MTAPSP model in terms of binary prediction performance, three validation methods were used in this study based on relying only on students' clickstream data and assessment score information:

1. The WF-PD task, which categorizes "Distinction" and "Pass" as one category and "Fail" and "Withdrawn" as another. The purpose of the WF-PD task is to identify whether a student has passed the course.
2. The F-PD task, which identifies students at risk of "Fail" by placing "Fail" in a separate category and "Distinction" and "Pass" in the same category, excluding dropouts. "The F-PD task identifies students at risk of "Fail".
3. The W-PD task, with "Withdrawn" in a separate category and "Distinction" and "Pass" combined in one category, excluding "Fail" students. "Fail" students were excluded, to distinguish between students at risk of dropping out and those not at risk.

To demonstrate the early risk prediction capability of the MTAPSP model, this paper divided the dataset by percentage of course length and created two models to predict student performance: (1) the "daily" model using student activity data as input daily; (2) the "weekly" model using students' weekly activity data as input. A comparison was made with the latest student performance prediction model, TEnc [24], and the results of the comparison are shown in Table 6.

Table 6. Comparison of prediction performance for model binary classification.

Models		Stages and Performance (%)									
		20%		40%		60%		80%		100%	
		Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**	**Acc**	**F1**
TEnc	WF-PD (daily)	75.41	71.50	82.83	80.91	88.23	86.77	90.78	89.93	93.42	93.01
MTAPSP	WF-PD (daily)	91.79	91.80	92.50	92.50	92.70	92.70	93.51	93.51	94.63	94.63
TEnc	F-PD (daily)	77.80	52.50	82.95	67.70	87.04	75.92	89.29	81.04	91.85	86.22
MTAPSP	F-PD (daily)	91.38	91.18	91.40	91.20	91.53	91.32	91.87	91.78	92.71	92.55
TEnc	W-PD (daily)	83.17	68.03	90.49	83.84	94.46	90.77	96.78	94.80	98.02	96.88
MTAPSP	W-PD (daily)	95.73	95.71	96.08	96.07	97.28	97.27	98.08	98.08	99.08	99.08
TEnc	WF-PD (weekly)	76.76	73.37	83.93	81.87	88.96	87.62	91.81	91.10	93.57	93.21
MTAPSP	WF-PD (weekly)	91.02	91.03	91.99	91.99	92.59	92.59	93.17	93.18	94.32	94.33
TEnc	F-PD (weekly)	78.53	55.68	83.73	69.20	87.85	77.41	90.43	83.09	91.75	86.27
MTAPSP	F-PD (weekly)	91.00	90.77	91.17	91.00	91.15	91.01	91.95	91.83	92.71	92.54
TEnc	W-PD (weekly)	83.79	70.63	91.01	84.79	95.51	92.61	97.22	95.58	98.14	97.10
MTAPSP	W-PD (weekly)	94.46	94.43	96.10	96.09	96.85	96.84	97.94	97.94	98.92	98.92

An analysis of Table 6 reveals the following:

1. The proposed method in this paper achieved optimal performance across various course durations, demonstrating the superior effectiveness of MTAPSP in binary predicting student performance.
2. Compared with the TEnc model, the three-layer LSTM network owned by MTAPSP performs better in dealing with long sequence data, indicating that the three-layer LSTM is more adept at capturing the patterns of long sequence data.
3. Within the first 20% of the course start time, MTAPSP embodies a high prediction performance, which indicates that the ANN in the MTAPSP model can more accurately map the relationship between the assessment scores characteristics and the students' final performance and map the potential influence of the assessment scores characteristics on the students' final performance.
4. Numerically, MTAPSP's daily model performs better than the weekly model, with performance improvements ranging from 0.1% to 1%, which suggests that the aggregation of student activity data changes important feature values, resulting in MTAPSP's inability to learn better features from this aggregated data.

To further demonstrate the early prediction ability of MTAPSP, the prediction performance of the MTAPSP model was compared with that of the TEnc model over three months after the start of the course, and the results are shown in Figures 5 and 6.

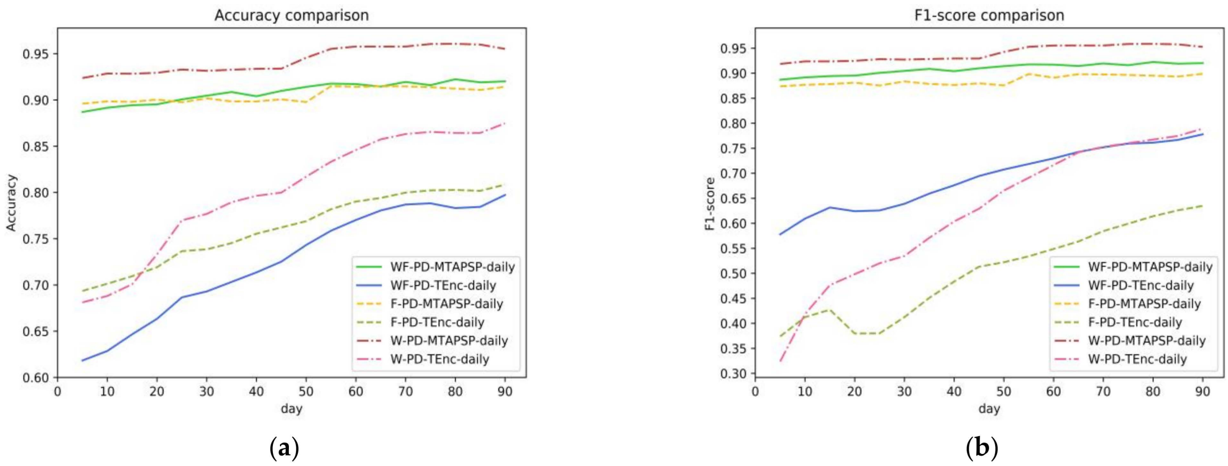
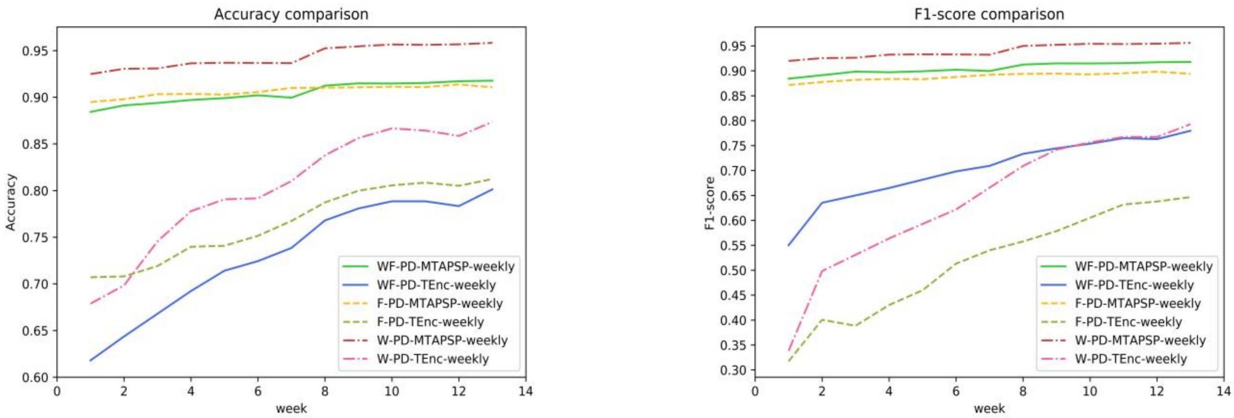


Figure 5. (**a**) Comparison of the accuracy of "daily" models based on the WF-PD task, based on the F-PD task, and based on the W-PD task three months after the start of the course; (**b**) comparison of the F1-scores of "daily" models based on the WF-PD task, based on the F-PD task, and based on the W-PD task three months after the start of the course.

As can be seen from Figures 5 and 6, in the WF-PD, F-PD, and W-PD tasks, at the beginning of the course, the daily and weekly models of MTAPSP achieve more than 85% in terms of accuracy and recall in predicting students' performance, and gradually increase the indexes over time, outperforming the TEnc model.

In the early risk prediction experiment, the MTAPSP model demonstrated excellent early prediction capabilities, enabling instructors to predict final student performance early in the course. Once the model predicts that a student is likely to end up with a "fail" performance, the instructor can develop an intervention plan to support these students to adjust their learning before the end of the course to ensure that they can successfully pass the exam. With the predictive results of the MTAPSP model, teachers can identify potential student problems ahead of time and help them overcome their difficulties through effective measures, thus improving overall student performance and learning experience. This early prediction and intervention approach helps to maximize the optimization of the student learning process and improve teaching and learning outcomes.



(**a**) (**b**)

Figure 6. (**a**) Comparison of the accuracy of "weekly" models based on the WF-PD task, based on the F-PD task, and based on the W-PD task three months after the start of the course; (**b**) comparison of the F1-scores of "weekly" models based on the WF-PD task, based on the F-PD task, and based on the W-PD task three months after the start of the course.

5.3.3. Ablation Experiments

To verify the effect of different modules on MTAPSP, ablation experiments were performed on MTAPSP models. Where MTAPSP-M-A denotes the MTAPSP model after removing MHSA and ANN, MTAPSP-A denotes the MTAPSP model after removing ANN, and the MTAPSP-M model denotes the MTAPSP model after removing MHSA. The experimental results are shown in Table 7.

Table 7. Results of ablation experiments.

Model	Precision	Recall	F1-Score	Accuracy
MTAPSP-M-A	0.643	0.641	0.626	0.703
MTAPSP-A	0.650	0.588	0.594	0.719
MTAPSP-M	0.697	0.576	0.572	0.723
MTAPSP	0.692	0.656	0.668	0.744

Figure 7a–d shows the confusion matrix of the predictive model at the time of the ablation experiment, with the diagonal line being the number of correctly predicted categories. Combined with Table 7, it can be seen that compared to MTAPSP-M-A, MTAPSP-A's accuracy and precision are increased, but its recall and F1 score are decreased, a phenomenon that stems from the problem of imbalance in the distribution of the sample categories in the dataset. As shown in Figure 7b, MTAPSP-A is inferior to MTAPSP-M-A in identifying the categories "Distinction" and "Withdrawn", while the number of correct identifications is significantly higher than that of MTAPSP-M-A in the categories "Fail" and "Pass", and "Pass" is significantly higher than in MTAPSP-M-A. Although the overall accuracy increased, the overall recall and F1 score decreased due to the decrease in the performance of the metrics in the other categories. The significant difference in precision and accuracy of the MTAPSP-M model compared to the MTAPSP-M-A model is because the ANN can fit arbitrary nonlinear functions, which allows it to learn the assessment scores and demographic features well and improve the model performance. Compared with MTAPSP-M-A, the overall number of correctly predicted categories of MTAPSP is substantially increased and the overall performance metrics are significantly improved. This is because MHSA can focus on different dimensions of features such as student learning behavior, assessment scores, and demographic information, and, when paired with ANN, can better integrate student learning behavior features, assessment scores, and demographic features to further improve model performance.

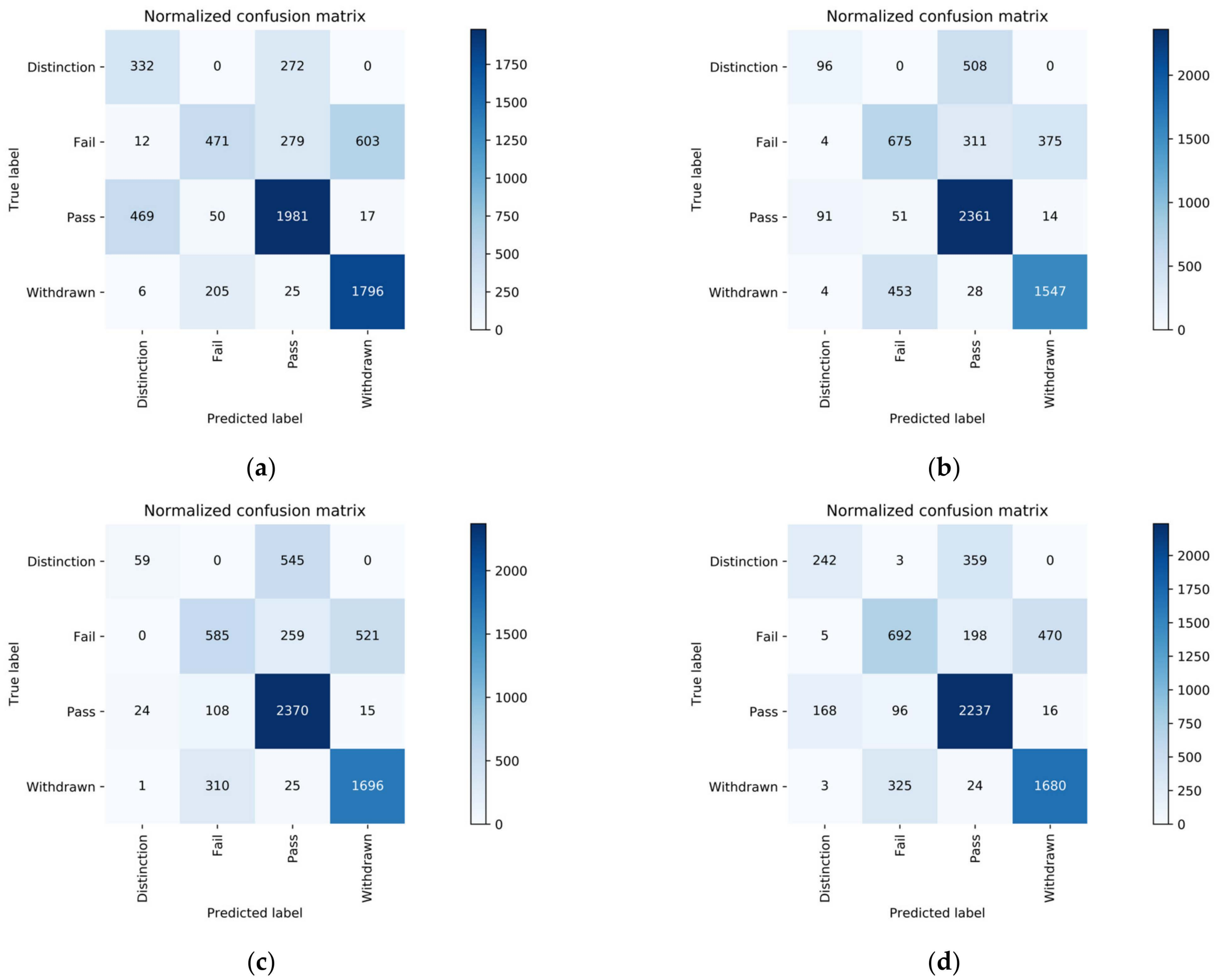


Figure 7. (**a**) Normalized confusion matrix for MTAPSP-M-A model predictions; (**b**) normalized confusion matrix for MTAPSP-A model predictions; (**c**) normalized confusion matrix for MTAPSP-M model predictions; (**d**) normalized confusion matrix for MTAPSP model predictions.

6. Conclusions

This paper proposes a student performance prediction model (MTAPSP) based on multidimensional time-series data analysis. The model incorporates multidimensional data such as students' learning behaviors, assessment scores, and demographic information (e.g., age group, place of residence) to achieve the multi-classification prediction of students' performance, and the multi-classification prediction performance of the MTAPSP model is verified by comparison with the baseline model. In addition, this paper tests the model's binary early prediction ability by dividing the dataset by course duration, and the experimental results show that it exhibits excellent performance in multi-classification prediction and binary early prediction. In future research, there is a need to address the imbalance in the data sample and to explore the impact of student–teacher interaction and student-to-student interaction data on student academic performance.

Author Contributions: Conceptualization, Z.S. and M.X.; methodology, M.X.; software, M.X.; validation, M.X., J.M. and H.Z.; formal analysis, H.Z.; investigation, J.M.; resources, Z.S.; data curation, M.X.; writing—original draft preparation, M.X.; writing—review and editing, Z.S.; visualization, M.X.; supervision, Z.S.; project administration, Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (62177012, 61967005, 62267003) and was supported by the Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory (GXKL06240107).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alhothali, A.; Albsisi, M.; Assalahi, H.; Aldosemani, T. Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability* **2022**, *14*, 6199. [CrossRef]
2. Tao, T.; Sun, C.; Wu, Z.; Yang, J.; Wang, J. Deep Neural Network-Based Prediction and Early Warning of Student Grades and Recommendations for Similar Learning Approaches. *Appl. Sci.* **2022**, *12*, 7733. [CrossRef]
3. Hughes, G.; Dobbins, C. The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). *Res. Pract. Technol. Enhanc. Learn.* **2015**, *10*, 10. [CrossRef]
4. Aljohani, N.R.; Fayoumi, A.; Hassan, S.U. Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability* **2019**, *11*, 7238. [CrossRef]
5. Liu, Y.; Fan, S.; Xu, S.; Sajjanhar, A.; Yeom, S.; Wei, Y. Predicting Student performance using clickstream data and machine learning. *Educ. Sci.* **2022**, *13*, 17. [CrossRef]
6. Casalino, G.; Castellano, G.; Zaza, G. Explainable Fuzzy Models for Learning Analytics. In *International Conference on Intelligent Systems Design and Applications*; Springer: Cham, Switzerland, 2022; pp. 394–403.
7. Arashpour, M.; Golafshani, E.M.; Parthiban, R.; Lamborn, J.; Kashani, A.; Li, H.; Farzanehfar, P. Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization. *Comput. Appl. Eng. Educ.* **2023**, *31*, 83–99. [CrossRef]
8. Hussain, S.; Khan, M.Q. Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Ann. Data Sci.* **2023**, *10*, 637–655. [CrossRef]
9. Conijn, R.; Snijders, C.; Kleingeld, A.; Matzat, U. Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Trans. Learn. Technol.* **2016**, *10*, 17–29. [CrossRef]
10. Meier, Y.; Xu, J.; Atan, O.; Van der Schaar, M. Predicting grades. *IEEE Trans. Signal Process.* **2015**, *64*, 959–972. [CrossRef]
11. Ahmad, M.S.; Asad, A.H.; Mohammed, A. A Machine Learning Based Approach for Student Performance Evaluation in Educational Data Mining. In Proceedings of the 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 26–27 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 187–192.
12. Zhang, W.; Huang, X.; Wang, S.; Shu, J.; Liu, H.; Chen, H. Student performance prediction via online learning behavior analytics. In Proceedings of the 2017 International Symposium on Educational Technology (ISET), Hong Kong, China, 27–29 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 153–157.
13. Al-Shehri, H.; Al-Qarni, A.; Al-Saati, L.; Batoaq, A.; Badukhen, H.; Alrashed, S.; Alhiyafi, J.; Olatunji, S.O. Student performance prediction using support vector machine and k-nearest neighbor. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 30 April–3 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–4.
14. Wang, C.; Wei, X.; Yang, A.; Zhang, H. Construction and Analysis of Discrete System Dynamic Modeling of Physical Education Teaching Mode Based on Decision Tree Algorithm. *Comput. Intell. Neurosci.* **2022**, *2022*, 2745146. [CrossRef] [PubMed]
15. Rao, G.M.; Kumar, P.K.K. Students Performance Prediction in Online Courses Using Machine Learning Algorithms. *United Int. J. Res. Technol* **2021**, *2*, 74–79.
16. Pek, R.Z.; Özyer, S.T.; Elhage, T.; Özyer, T.; Alhajj, R. The role of machine learning in identifying students at-risk and minimizing failure. *IEEE Access* **2022**, *11*, 1224–1243. [CrossRef]
17. Jawad, K.; Shah, M.A.; Tahir, M. Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability* **2022**, *14*, 14795. [CrossRef]
18. Bujang, S.D.A.; Selamat, A.; Ibrahim, R.; Krejcar, O.; Herrera-Viedma, E.; Fujita, H.; Ghani, N.A.M. Multiclass prediction model for student grade prediction using machine learning. *IEEE Access* **2021**, *9*, 95608–95621. [CrossRef]
19. Hung, J.L.; Wang, M.C.; Wang, S.; Abdelrasoul, M.; Li, Y.; He, W. Identifying at-risk students for early interventions—A time-series clustering approach. *IEEE Trans. Emerg. Top. Comput.* **2015**, *5*, 45–55. [CrossRef]
20. Yu, C.C.; Wu, Y. Early warning system for online stem learning—A slimmer approach using recurrent neural networks. *Sustainability* **2021**, *13*, 12461. [CrossRef]
21. He, Y.; Chen, R.; Li, X.; Hao, C.; Liu, S.; Zhang, G.; Jiang, B. Online at-risk student identification using RNN-GRU joint neural networks. *Information* **2020**, *11*, 474. [CrossRef]
22. Liu, T.; Wang, C.; Chang, L.; Gu, T. Predicting High-Risk Students Using Learning Behavior. *Mathematics* **2022**, *10*, 2483. [CrossRef]
23. Qu, S.; Li, K.; Wu, B.; Zhang, S.; Wang, Y. Predicting student achievement based on temporal learning behavior in MOOCs. *Appl. Sci.* **2019**, *9*, 5539. [CrossRef]

24. Kusumawardani, S.S.; Alfarozi, S.A.I. Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities. *IEEE Access* **2023**, *11*, 18960–18971. [CrossRef]
25. Chen, H.-C.; Prasetyo, E.; Tseng, S.-S.; Putra, K.T.; Prayitno; Kusumawardani, S.S.; Weng, C.-E. Week-Wise Student Performance Early Prediction in Virtual Learning Environment Using a Deep Explainable Artificial Intelligence. *Appl. Sci.* **2022**, *12*, 1885. [CrossRef]
26. Song, X.; Li, J.; Sun, S.; Yin, H.; Dawson, P.; Doss, R.R.M. SEPN: A sequential engagement based academic performance prediction model. *IEEE Intell. Syst.* **2020**, *36*, 46–53. [CrossRef]
27. Aljaloud, A.S.; Uliyan, D.M.; Alkhalil, A.; Elrhman, M.A.; Alogali, A.F.M.; Altameemi, Y.M.; Altamimi, M.; Kwan, P. A deep learning model to predict Student learning outcomes in LMS using CNN and LSTM. *IEEE Access* **2022**, *10*, 85255–85265. [CrossRef]
28. Chen, Y.; Wei, G.; Liu, J.; Chen, Y.; Zheng, Q.; Tian, F.; Zhu, H.; Wang, Q.; Wu, Y. A prediction model of student performance based on self-attention mechanism. *Knowl. Inf. Syst.* **2023**, *65*, 733–758. [CrossRef]
29. Li, M.; Wang, X.; Wang, Y.; Chen, Y.; Chen, Y. Study-GNN: A novel pipeline for student performance prediction based on multi-topology graph neural networks. *Sustainability* **2022**, *14*, 7965. [CrossRef]
30. Yang, T.-Y.; Brinton, C.G.; Joe-Wong, C.; Chiang, M. Behavior-based grade prediction for MOOCs via time series neural networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 716–728. [CrossRef]
31. Kuzilek, J.; Hlosta, M.; Zdrahal, Z. Open university learning analytics dataset. *Sci. Data* **2017**, *4*, 170171. [CrossRef] [PubMed]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, N. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
33. Alsalman, Y.S.; Halemah NK, A.; AlNagi, E.S.; Salameh, W. Using decision tree and artificial neural network to predict students academic performance. In Proceedings of the 2019 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 11–13 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 104–109.
34. Adnan, M.; Habib, A.; Ashraf, J.; Mussadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access* **2021**, *9*, 7519–7539. [CrossRef]