Article

A Study on the Learning Early Warning Prediction Based on Homework Habits: Towards Intelligent Sustainable Evaluation for Higher Education

Wenkan Wen [1] , Yiwen Liu [1,2,3,*], Zhirong Zhu [1] and Yuanquan Shi [1,2,3]

1   School of Computer and Artificial Intelligence, Huaihua University, Huaihua 418000, China
2   Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, Huaihua 418000, China
3   Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaihua 418000, China
*   Correspondence: lyw@hhtc.edu.cn; Tel.: +86-15367585173

Abstract: Teachers need a technique to efficiently understand the learning effects of their students. Early warning prediction mechanisms constitute one solution for assisting teachers in changing their teaching strategies by providing a long-term process for assessing each student's learning status. However, current methods of building models necessitate an excessive amount of data, which is not conducive to the final effect of the model, and it is difficult to collect enough information. In this paper, we use educational data mining techniques to analyze students' homework data and propose an algorithm to extract the three main features: Degree of reliability, degree of enthusiasm, and degree of procrastination. Building a predictive model based on homework habits can provide an individualized evaluation of students' sustainability processes and support teachers in adjusting their teaching strategies. This was cross-validated using multiple machine learning algorithms, of which the highest accuracy was 93.34%.

Keywords: sustainable evaluation; educational prediction; early warning; homework habits; machine learning

1. Introduction

1.1. Background

Institutions of higher education have altered their educational environments and practices to support online instruction in response to the increasing number of new COVID-19 cases and school closures, and the virus has forced the implementation of home quarantine to prevent the spread of the disease [1]. Online remote teaching and learning solutions have emerged as a critical crisis response to ensure that teaching and learning can proceed. Online learning has become a very common method of teaching and learning because, according to UNESCO, the majority of nations and regions use digital distance learning to maintain the stability and continuity of education in the face of the educational stagnation brought on by the COVID-19 outbreak.

The online learning environment is a different mode of interaction, and much research attention has focused on how this new medium can facilitate learning in different ways and at different scales [2]. The way teachers teach in traditional classrooms is different from the way they teach online [3], and it is difficult for teachers to directly observe students' learning, which may not always accurately reflect the true state of students' learning if evaluated directly and roughly through basic factors such as homework completion, daily attendance, and classroom quizzes. At the same time, the lack of engagement and insufficient interaction between teachers, students, and content may have an impact on students' online learning, as not all students are satisfied with the online learning

environment, and the fact that students do not always have access to sufficient learning materials online also has an impact [4,5]. In addition, teachers' perceptions of the learning environment also affect their teaching methods [6], as teachers are not effectively informed about students' academic status and current learning environment and therefore do not receive timely and effective feedback to change their teaching responses. This results in a lack of effective student–teacher interactions and makes teaching organization and management difficult.

1.2. Application and Shortcomings of Student Portrait

Student portraits are a new method to analyze the basic information and learning behavior of online learners. By collecting multiple pieces of information and using data mining to build student portraits for personalized contextualization and learning process guidance, it can positively contribute to online learning and effectively help solve the problems mentioned above [7,8]. Student portraits are used to understand the motivation, goals, and learning behaviors of a highly diverse student population, and each student's student portrait is unique and can truly reflect their learning status and be used to propose targeted solutions to improve their learning [9]. We can not only help teachers change their teaching strategies through a sustainable process of evaluation of each student's learning status by creating student portraits, but we can also personalize education by pushing personalized learning resources based on students' learning needs [10].

However, there are many problems with the existing methods of constructing student portraits. First, the issue of collecting the information required to create student portraits covers all facets of students' everyday behavioral lives, which are challenging to gather and complicated, necessitating a significant number of people and resources [11]. Second, the most important issue in educational data mining is "privacy," i.e., privacy issues arising from the collection, analysis, dissemination, and use of personally identifiable data [12,13]. The education sector should avoid compiling excessive amounts of data on students, such as their Internet usage, book-borrowing habits, and spending patterns, as this may make them fear school [14]. As a result, it is challenging to gather extensive student data to create a student profile. However, from the perspective of homework, it is a novel idea to analyze homework record data to determine students' learning habits and produce tailored student reports to help teachers improve their lesson plans. The relationship between homework and academic accomplishment is crucial because it connects teaching and learning [15]. Most students will use online platforms to finish their homework, especially in the present pandemic climate, which also creates opportunities for data collection.

1.3. The Impact of Study Habits on Students

Learning habits are a way or pattern of behavior that recurs in the learning process and can be measured by the learner's learning behavior, which has a certain degree of plasticity and measurability [16,17]. The motivational factors that influence learning habits include internal motivational factors such as motivation, cognitive ability, and personality, as well as external motivational factors such as the task and environment; correlating these factors with poor learning habits, key factors for the emergence of poor learning habits can be identified, and the occurrence of poor learning behaviors can be reduced and the occurrence of target learning behaviors increased through interventions [18]. The FBM behavioral model states that three elements must be present simultaneously for a behavior to be achieved, namely, sufficient motivation, the ability to perform this behavior, and the trigger that causes the behavior to occur, and all three elements must be present simultaneously for the behavior to occur [19]. Therefore, how to construct characteristics from study habits as a way to grasp students' ability levels, motivational strengths, and triggers for changing learning states is of paramount importance.

Barbara Flunger [20] found that when students spend a lot of time on homework, they have low motivation and responsibility; by analyzing students' homework data, it was found that excessive homework time does not necessarily determine that learners

have good (or bad) signs of motivation. Not only that, students' completion of homework before the deadline and academic procrastination are common phenomena, and academic procrastination largely reflects the manifestation of students' low motivation to learn. A common poor study habit, the academic procrastination habit is the delayed behavior presented when learners are faced with the stimulus of an event related to homework. A previous study showed that [21] students who exhibit a higher tendency to procrastinate obtain lower grades than students with a lower tendency to procrastinate. Academic procrastination habits also affect learners' performance, hinder their academic progress, increase their stress, and reduce the quality of their lives, negatively affecting their physical and mental health [22]. Furthermore, the level of motivation of students towards homework is closely related to their motivation to learn; Xu, Jianzhong [23] found that it is difficult to motivate students to approach homework enthusiastically; students who are strong learners may find homework boring and unchallenging; at the other extreme, students who are weak learners may find it laborious and torturous. Motivation to do homework depends heavily on student characteristics, and how to effectively target different learning programs to students of different ability levels is the key to motivation. Several studies have demonstrated [24–27] that the most common dishonest activity is cheating on homework, with rates as high as 45% of those surveyed. It is incredibly simple to duplicate and modify electronic documents to commit homework plagiarism, especially with the rise in online education. Plagiarism can be used to accomplish homework tasks quickly and with a high grade. Due to the outdated nature of the majority of the homework given by Chinese professors and the ease with which the solutions can be found online, the prevalence of plagiarism is much higher in China. This not only seriously undermines students' ability and learning initiative but also prevents teachers from making accurate assessments of students' true levels and obtaining effective teaching feedback from students' homework. This has led to a lack of interaction between teachers and students, which has led to a failure of teaching and a serious learning crisis.

Although a large number of related scholars have conducted relevant research on student portraits and achieved certain results, there are still many problems. For example, the lack of process-oriented evaluation, the difficulty of data collection, and the general recognition effect [28–31] have resulted in a single evaluation effect and thus prevented students and teachers from obtaining effective feedback. The present study aims to use educational data mining techniques to analyze student homework data, extract homework habit features using the proposed algorithm, and create a model based on learning habits. The model will be used to evaluate students over time and assist teachers in developing their lesson plans.

2. Framework Design

This study uses educational data mining techniques to extract basic student learning data through data cleaning and data transformation, followed by using proposed algorithms to analyze students' homework habits and construct predictive models. The models can assess students' academic levels and identify potential strengths, help identify students who are potentially at risk, assist teachers in lesson planning, and provide personalized learning analysis tailored to each student.

The proposed method is divided into four parts: (1) Collecting student log information on the learning platform, cleaning and transforming the data, and using data mining techniques to obtain the required data; (2) extracting features of each student's homework habits using the proposed algorithm; (3) constructing a prediction model based on the extracted features; and (4) based on the prediction results, generating an achievable, personalized analysis and assisting the teacher in making improvements to the teaching plans. The flowchart of prediction is shown in Figure 1.
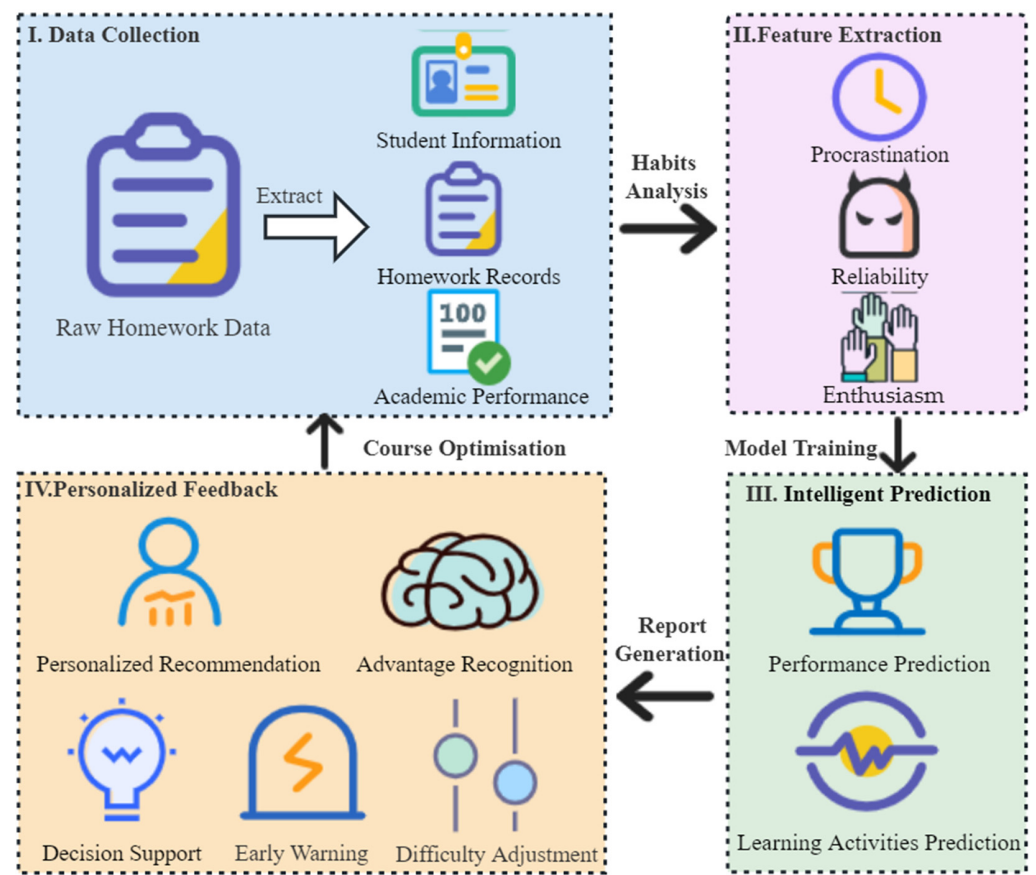
Figure 1. Flow chart of the proposed method. The method will be divided into four parts. (**I**) Collection, cleaning, and transformation of the data. (**II**) Extraction of features using the proposed algorithm. (**III**) Construction of the model. (**IV**) Generating personalized reports based on the results.

The dataset contains a total of 15 attributes, and several methods of logistic regression, decision trees, support vector machines, and CatBoost were used for model building. This study used 10-fold cross-validation to create an optimal method for evaluating model performance and assessed the predictive accuracy of the model by measuring accuracy, precision, recall, and F1 scores.

3. Data Collection

Directly relevant characteristics (such as test and quiz scores) are frequently employed for the early prediction of overall academic accomplishments, such as graduation credits or final grade ratings. The assessment can be used to predict student performance early, but the predictive effect is somewhat lacking. It is possible to predict student learning status more accurately by gathering underlying student homework habits. These relevant attributes are highly correlated with the data collected and predicted, enabling a more effective prediction of students' academic performance.

A large amount of learning log data were generated by the students' logs, such as daily attendance and homework completion level. By collecting this log data and using data mining techniques to archive and filter the data, we ended up with several dimensions of information. The data collected are divided into three categories: (1) Student information (gender, professional Information, etc.), (2) homework records (homework submission time, number of homework tasks submitted, etc.), (3) academic performance (homework grade ranking, total performance, etc.). The final scores were used as the final indicators to evaluate the students and as labels for the model predictions and the results are shown in Table 1.

Table 1. Statistics of students' learning records.

Number	Item	Describe
1	Gender	Gender
2	Age	Age
3	Types of Homework Questions	Multiple choice questions, programming questions.
4	Professional Information	Undergraduate major
5	Homework Submission Time	Time for students to submit homework
6	Number of Homework Submissions	Number of homework submitted by students
7	Homework Completion	Students complete their homework or not
8	Homework Grade Ranking	Ranking of students' homework grades
9	Change of Homework Ranking	The number of the (last- current) ranking
10	Course Video Completion Degree	Completion rate of students watching the video
11	Daily Attendance	Number of daily attendances
12	Final Score	Course grades (including regular grades and final grades)

4. Feature Extraction

4.1. Variable Analysis

Student performance and homework practices are closely related. This study examines three factors: Students' confidence in completing homework independently, students' motivation to complete homework, and students' habit of procrastinating on homework.

4.1.1. Degree of Reliability

Online homework is usually submitted as electronic documents or completed as online questions, and students can easily copy and modify electronic documents through development tools and electronic document editors, making it simple to complete homework in a very short time as a way to obtain high scores [32–34]. Therefore, we assess the validity of students' homework completion in our study using variables such as homework time and performance. The degree to which it is credible that students performed their homework independently is indicated by the reliability score.

Firstly, consider that in the process of completing homework, students are likely to have to interrupt their homework due to several things. As a result, there are situations where homework cannot be completed at one time but may have to be completed on several occasions. A key focus is then on how to effectively calculate the true homework time for students. The easiest way to think of it is to use the method of subtracting the last submission time from the initial submission time. However, this would calculate a very large homework time, and it would be difficult to distinguish the true homework time for each learner. Therefore, for the calculation of homework time, we use the interval between submission times as a basis for whether students consistently complete their homework and whether they do not have a too-long interval between two homework answer submission time points when working on the online platform.

It is only through the correct calculation of the time spent on the homework that it is possible to find that for some students who are given high grades for very short homework, it is likely that they did not complete the homework independently but rather by copying or searching for answers to the homework. However, it is not possible to judge whether a student has committed plagiarism based on time alone. We believe that students who score below a certain level on their homework performance cannot be judged to have plagiarized, because if a student does not continue with the rest of the homework after completing only part of the topic, the calculated time for the homework may also be short, but the student has not plagiarized.

4.1.2. Degree of Enthusiasm

The type of homework assigned by teachers has an important connection to students' motivation to learn; if the homework assigned attracts their attention, they will do it with the purpose of learning. However, if the homework does not attract their attention, they

will do the homework, and that is only because they are obliged to [35]. Therefore, via the degree of motivation, we can determine the level of interest of students in a particular course. Students who are interested in the course are more likely to complete the homework as soon as possible, while those who are not interested are likely to complete the homework mechanically by the deadline or simply forget about the class.

4.1.3. Degree of Procrastination

Academic procrastination habit is a common poor study habit, and the academic procrastination habit refers to the delayed behavior presented when learners are faced with the stimulus of an event related to homework. Studies have shown that students who show a higher tendency to procrastinate obtain lower grades than students with a lower tendency to procrastinate [36]. It is important, especially for teachers, to be aware of students' behaviors, especially their tendency to procrastinate. Since academic procrastination habits have a direct impact on academic performance, they are also a concern for many researchers and teachers [37].

This study proposes using students' homework start time, homework completion times, and homework duration to represent students' procrastination on homework. Homework procrastination is a measure of student procrastination, and the higher the procrastination, the more difficult it is for students to develop good study habits, and the higher achieving students generally have lower levels of procrastination. A previous article [16] used the number of times beyond the required time to do homework and the number of times beyond the required time to submit homework to determine students' procrastination, but most homework has deadlines and cannot be performed after this deadline. Therefore, this paper inferred students' procrastination level by comparing the average homework time, which is not easily influenced by the difficulty of the homework and can be adapted to a wider range of courses.

In Table 2, the factors influencing the degree of reliability, degree of enthusiasm, and degree of procrastination are discussed and analyzed several times, with a detailed analysis of the relevant influencing factors.

Table 2. Table of factors related to homework habits.

Items	Degree of Reliability	Degree of Enthusiasm	Degree of Procrastination
Homework Time	✓		✓
Homework Average Time	✓		✓
Homework Difficulty	✓		
Homework Deadline		✓	
Homework Interval			✓
Homework Score	✓		
Homework Average Score	✓		
Number of Homework Submissions			✓
First Submission Time		✓	
Last Submission Time		✓	

4.2. Algorithm Design

The method of extracting work habits was divided into three stages, which ultimately returned three feature vectors: The degree of reliability, degree of enthusiasm, and degree of procrastination.

Algorithm 1 shows the detailed procedure of homework hours calculation. The algorithm requires the input of three values, ω, δ and λ which are the homework submission record, the homework interval threshold, and the homework difficulty factor, respectively.

For each student, a value of time is calculated for each homework assignment. Here, *i* is the homework timestamp and *n* is the total number of homework submission records. The algorithm returns the homework duration.

Algorithm 1 Calculation of homework time

Require: ω, δ, λ
While *i* < *n*
   If $(\omega_{i+1} - \omega_i) < \delta$ **then**
     time + = $(\omega_{i+1} - \omega_i)$
   End if
   If *i* + 1 = *n* **then**
     Return *time*/λ
End while

ω_i, ω_{i+1}, respectively, are the time stamps for the submission of the two homework answers, that is, $(\omega_{i+1} - \omega_i)$ is the interval between the two submission times. δ is the set threshold. If the interval between two submission times is $(\omega_{i+1} - \omega_i) > \delta$ then it is not considered to be continuous homework. Otherwise, it will be considered continuous working time. The difference between the two submission times is used as the operation time, and the operation time of several times is accumulated to obtain the current operation time. This is *time* as the homework time and λ as the homework difficulty factor, of which the default is 1. At that time, *i* = *n* it means that all records have been traversed, and the operation time will be *time* divided by the operation difficulty factor λ to obtain the weighted operation time. Among them ω_i, ω_{i+1}, $\delta \in N$, *i* = *i* ... *n*, *n* are positive integers, $\lambda \in (0,2)$.

Algorithm 2 is the detailed process of plagiarism calculation, *C*. The algorithm needs to input multiple values, *P*, P_{ave}, *time*, $time_{ave}$, ε and θ, which represent the homework score, homework average score, homework time, average homework time, plagiarism score factor, and plagiarism time factor. For each student and *C* the value calculated for each homework, the algorithm returns the number of plagiarisms *C*. In order to be careful, first of all, *time* the average homework time is calculated by adding up the homework time of all students $time_{ave}$, as shown in Equation (1). Then the average homework time is calculated by adding up P_{ave} the homework time of all students *P*, as shown in Equation (2). If the homework score *P* is greater than the product P_{ave}, of ε the sum, a subsequent judgment will be made, and if it is less than *C*, the value will not be changed. In the subsequent judgment, if the working time *time* is less than the product $time_{ave}$ of θ the sum, it will be judged as plagiarism, and *C* the value will be increased by one. Among them, *i* = *i* ... *n*, *n* is the total number of students, $\varepsilon, \theta \in (0,1)$.

$$time_{ave} = \frac{1}{n}\sum\nolimits_{i=1}^{n} time_i \tag{1}$$

$$P_{ave} = \frac{1}{n}\sum\nolimits_{i=1}^{n} P_i \tag{2}$$

Algorithm 2 Judgement of plagiarism

Require: *P*, P_{ave}, *time*, $time_{ave}$, ε, θ
If $P < P_{ave}{}^* \times \varepsilon$ **then**
     C + = 0
Else
    If *time* < $time_{ave} \times \theta$ * θ
     C + = 1
    End if
End if

Algorithm 3 feature vectors x_1, x_2, are x_3 are shown in the detailed process of calculation. The algorithm needs to input multiple values, namely C, D, $time_w$, F_s, ω, $time$, and $time_{ave}$, δ, which are the number of plagiarisms, the homework deadline, the first submission time, the judgment factor, the homework time, the average homework time, and the delay factor, where the total number of students is n. The algorithm returns the feature vectors x_1, x_2, x_3.

Algorithm 3 Calculation of homework habits

Require: *C*, *D*, $time_w$, F_s, ω, $time$, $time_{ave}$, δ
While $i < n$
$x_1 < -\frac{C_i}{n}$
  If $F_{si} - D_i < time_{wi} \times \omega$ **then**
    $x_2 += 1$
  Else
    If $F_{si} - D_i > time_{wi} \times (1 - \omega)$ **then**
      $x_2 -= 1$
    Else
      $x_2 += 0$
    End if
  End if
  If $time_i < (1+\delta) \times time_{avei}$ **then**
    $x_3 += 1$
  End if
End while

First, the value of the feature vector is obtained by the ratio of the number of plagiarisms to the total number of x_1 homework assignments. Algorithm 2 can calculate the total number of times the learner is judged to be plagiarizing homework. If the difference between the first submission time F_s and the homework posting time is less than D the value of the product of the x_2 homework deadline $time_w$ and the judgment factor, it will be ω increased by one. Otherwise, if it is greater than $time_w$ the value of the product of the homework deadline x_2 and $(1 - \omega)$, it will be ω decreased by one, otherwise no x_2 change will be made. Furthermore, if the work time $time$ is less than the average work $time_{ave}$ the value of the product of the $(1 + \delta)$ delay factor, x_3 we add one to the value.

5. Research Procedure

5.1. Data Source

The original data used in our study was obtained from the Principles of Computer Composition course on the Programming Teaching Assistant (PTA) platform in the Spring 2022 semester of the second year of undergraduate studies at Huaihua University, and all data for this course were authorized for use by the university. The course includes online exercises, video lessons, quizzes, and other exercises and resources, while mid-term and final exams are required for credit and the final grade is graded by several professors, with a final grade of >70 required to be deemed a pass.

5.2. Data Statistics

There were 161 learners in the course, including 29% female and 71% male learners, all around the age of 21 (M = 21.38, SD = 0.95). This course was chosen for the following reasons: (1) Students were more active and more willing to use the online platform to learn, which provided a wealth of valuable behavioral data, and (2) compared to other courses, the selected course has more students.

This study was considered a classification problem and therefore students were classified into three categories, namely, excellent, good, and poor based on their final scores, and the results are shown in Table 3.

Table 3. Final score statistics table.

Grade	Describe	Percentage	Mean/SD
Excellent	$90 \leq$ Score ≤ 100	10.69%	93.83/3.29
Good	$70 \leq$ Score < 90	71.07%	67.82/7.84
Poor	$0 <$ Score < 70	18.24%	32.14/19.12

All the data collected from student work records were subjected to data cleaning and transformation operations. In this study, the SHAP value was used for attribute selection to improve the accuracy and efficiency of data prediction [38]. Some unnecessary attributes were removed (e.g., question type, ID, and student name), and the appropriate attributes were finally selected. Details are shown in Table 4.

Table 4. Dataset attributes and contribution ranking.

Number	Attribute	Content	Average Merit	Mean/SD
1	m_1	Score for the first homework	0.005	81.51/46.52
2	m_2	Score for the second homework	0.043	80.42/46.65
3	m_3	Score for the third homework	0.016	80.18/46.20
4	m_4	Score for the fourth homework	0.081	79.87/45.99
5	m_5	Score for the fifth homework	0.089	80.66/46.71
6	t_1	Time for the first homework (Second)	0.021	3029.92/1744.35
7	t_2	Time for the second homework (Second)	0.066	7906.62/5125.38
8	t_3	Time for the third homework (Second)	0.001	10,022.33/6046.22
9	t_4	Time for the fourth homework (Second)	0.053	5621.80/4136.32
10	t_5	Time for the fifth homework (Second)	0.088	8355.34/5289.48
11	x_1	Degree of Reliability	0.157	0.21/0.27
12	x_2	Degree of Enthusiasm	0.121	0.06/0.12
13	x_3	Degree of Procrastination	0.164	0.19/0.27
14	s	Score for daily student behaviors	0.095	69.92/45.18

There were a total of 5 homework records in this dataset. Among them, m_{i1} is the grade of a student's first homework and t_{i1} is the homework time of a student's first homework. The feature vectors x_{i1}, x_{i2} and x_{i3} are the degree of reliability, the degree of enthusiasm, and the degree of procrastination, respectively. The comprehensive score is represented by s_i. Therefore, a total of 14 attributes are used to represent each student in the course. The class attribute is used to determine the classification category. Therefore, the final dataset has a total of 15 variables (as shown in Equation (3)) as follows:

$$m_{i1}, m_{i2}, \ldots, m_{i5}, t_{i1}, t_{i2}, \ldots, t_{i5}, x_{i1}, x_{i2}, x_{i3}, s_i, class \tag{3}$$

In this experiment, four mature machine learning models, LR, SVM, DT, and Cat Boost, are used for cross-validation, that is, the same data are used for training, verification, and testing. The experiment uses 10-fold cross-validation to validate the model. First, the dataset is divided into a training dataset and a test dataset. Then the training dataset is split into K parts; $(K - 1)$ parts are used for training and 1 part is used for verification, and the performance of each model is recorded. Until each k-fold is used for validation, the model hyperparameters with the optimal score mean and standard value will be generated, and finally, the model performance will be evaluated on the test dataset.

5.3. Evaluation Metric

The experimental procedure uses accuracy (as in Equation (4)), precision (as in Equation (5)), recall (as in Equation (6)), and the comprehensive evaluation index (*F1-score*) (as in Equation (7)) to evaluate the performance of classifier algorithms (*TP*: Correct rate; *FP*: False-positive rate; *FN*: False-negative rate). The *F1-score* is widely used in

information retrieval, machine learning, sentiment analysis, and other fields involving binary classification.

$$Accuracy = \frac{TP}{S_i} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

5.4. Results

It can be seen that all methods were able to predict overall student performance to some extent, which proves the validity of the selected features. Figure 2 shows that CatBoost has good predictive performance for all three categories, while SVM also achieves better prediction results.
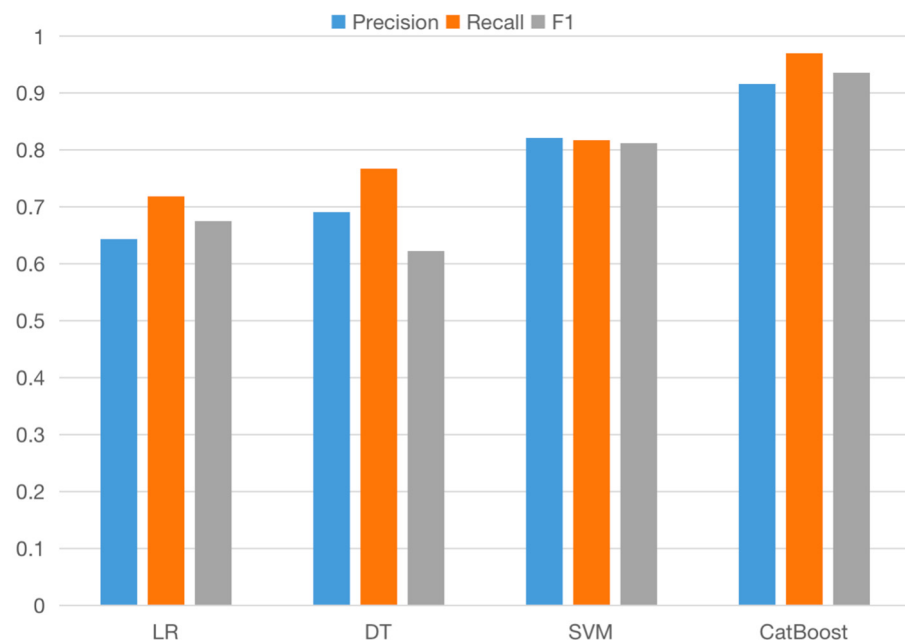
Figure 2. Classification metrics for each model.

As can be seen from Table 5, CatBoost has excellent performance, boasting a 93.34% accuracy, 91.67% precision, 96.97% recall, and 93.65% *F1*-score. CatBoost is an integrated learning algorithm with good predictive performance that can directly cascade different models. That is, the dataset is randomly sorted, and the average label value of samples with the same category values can be calculated by including only the label values of samples prior to this one. It was discovered that the suggested method could have a good predictive effect and that student work habits may accurately reflect students' study habits, accurately reflect their level of learning, and help people make wise decisions.

Table 5. Classification metrics table.

Model	Accuracy	Precision	Recall	F1
LR	74.23%	64.44%	71.88%	67.59%
DT	66.94%	69.17%	76.77%	62.34%
SVM	83.41%	82.19%	81.75%	81.25%
CatBoost	93.34%	91.67%	96.97%	93.65%

6. Discussion

6.1. Effectiveness of Homework Habits

For students, work habits have a large impact on attitudes toward work and interest in learning, as well as on the effectiveness of learners over time. Moreover, the work habits of students vary considerably between levels, allowing for more effective prediction.

Table 6 shows the average homework habit scores for students at each achievement level. As can be seen, students with high grades have a very high degree of reliability, and these students generally do not bother with plagiarism in their homework and are therefore able to achieve better results in their final exams. In contrast, students with poor grades are very likely to plagiarize. For students with good grades, we find that the degree of reliability is also high, but that they have a lesser degree of enthusiasm and a certain amount of procrastination compared to those with poor grades.

Table 6. Table of homework habit coefficients for each category of students.

Grade	Average Degree of Reliability	Average Degree of Enthusiasm	Average Degree of Procrastination
Excellent	0.996	0.823	0.013
Good	0.923	0.687	0.216
Bad	0.331	0.228	0.735

Balanced with this, students with a high degree of reliability, a high degree of enthusiasm, and a low degree of procrastination were generally better performers. This also suggests that homework habits can be an effective indicator of a student's level of learning.

6.2. Practical Implications

The significance of this research is that it develops a new model based on homework habits from the data generated by students' homework submission records on the online platform. This allows for the development of individualized learning plans for each student and the identification of at-risk students early in the learning process. At the same time, it provides teachers with decision support to better help them improve their teaching plans. In addition, students with different homework habits have different targeted solutions, which can better help students to solve their own problems.

A high level of plagiarism means that students often carry out plagiarized work and teachers target such students for intervention. For example, by assigning homework that cannot be plagiarized, such as video recordings, live answers, etc., we can effectively engage learners in active learning. Such students can also improve by using the same questions but different answers, even if they can find the homework answers on the Internet [39].

If they find that students are not motivated by the work, teachers should consider whether there are problems with their lesson plans. For example, is the amount of homework assigned too much? Is the homework assigned not educational and only expected to be completed within a limited time frame? Is the homework assigned perhaps not appropriate to the student's level and course content? Is it because the grading of the student's work is not reasonable? Teachers can assign different learning models to different levels of learners. Some high-achieving students may see homework as repetitive, ineffective, and not helpful to improve their learning. Other low achievers may feel that the homework is too difficult and takes a lot of time to complete, or they do not have access to other resources to help them better solve the homework. High-quality homework should be purposeful, appropriate to the content of the course, and at the student's level. Using differentiated instructions to solve problems, one should receive timely feedback during the teaching process, identify problems and solve them in a timely manner, and ask questions at the appropriate level for students of different levels and styles to see if the expected requirements are met.

For such students with high levels of procrastination, teachers should ask students whether they are procrastinating because of games, family, or deep-seated reasons. For students with lower levels of procrastination, teachers can target homework with shorter deadlines so that students are pressured to complete them as soon as possible. Academic procrastination is most likely due to a lack of confidence in their skill level and learning strategies rather than a lack of knowledge about them [40]. Teachers should also attempt to provide some level of encouragement to students with high levels of procrastination to help them regulate the occurrence of such states.

6.3. Personalized Feedback and Sustainability Assessment

By collecting student work records and analyzing student study habits, it is possible to identify potentially at-risk students and provide early warning without unduly invading their privacy. This analysis of the degree of reliability, the degree of enthusiasm, and the degree of procrastination enables teachers to dynamically vary the difficulty of students' homework, identify students' potential strengths, and develop personalized learning plans for students and continuously improve them. Each student will be able to access personalized analysis, understand their own shortcomings, and improve themselves through personalized recommendations.

For students, homework habits largely influence their attitudes and interest in homework and have a degree of influence on learning efficiency. Many students do not lack the ability to learn, but their current learning methods and study habits are not perfect. By analyzing the way students complete their homework, we can help them to change their own shortcomings, which will be of great benefit to their future lifelong learning and lifelong development. Learning is not something that can only be performed in school; learning is always a lifelong endeavor. Learning is not rote learning either, because learning is not just about learning knowledge, but rather using knowledge as a vehicle for students to learn what is known and to face the unknown.

Homework habits are significantly linked to human development, and study habits promote a person's lifelong development. As shown in Figure 3, students are helped to develop within the school through sustainable evaluation reports, teachers are given feedback through sound evaluations, and teachers optimize their own teaching programs to promote changes in students' study habits. This cycle helps students build good study habits. Good learning habits in turn help students' lifelong development, opening up the internal and external cycle to better achieve sustainable education.
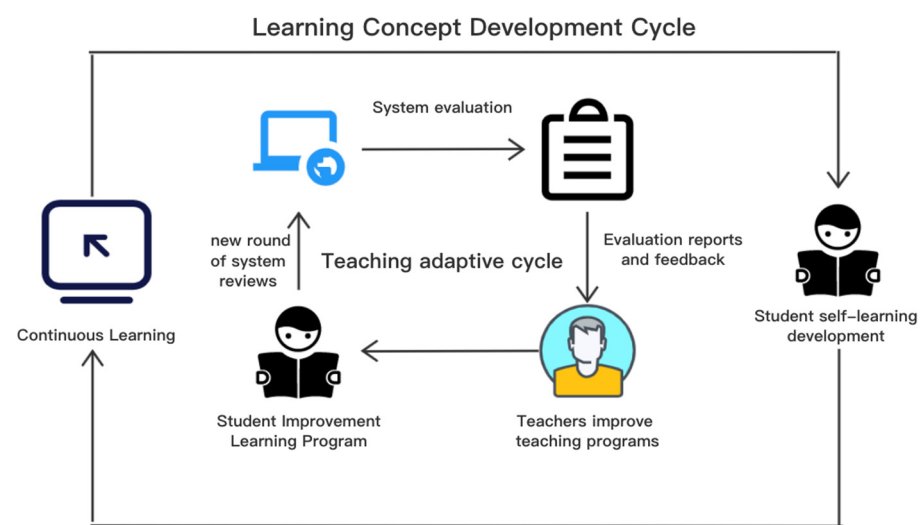

Figure 3. Sustainability evaluation cycle.

In addition, the use of artificial intelligence for teaching evaluation is smarter and more reasonable. The homework of the course is no longer limited to a summative evaluation

at the end of the course but rather adopts diversified homework indexes for the teaching objectives and constructs a process evaluation system to reflect the learners' mastery of the course to the maximum extent. The use of intelligent, sustainable AI to simplify the evaluation process and grasp the learning status of learners not only ensures fairness but also allows tutors, class teachers, and full-time teachers to intervene more effectively in advance [41]. At the same time, the use of AI for teaching evaluation can effectively cover every student, and each student can be given a personalized evaluation report with targeted advice to effectively stimulate students' enthusiasm, interest, and motivation for learning [42–46].

7. Conclusions

In this study, we endeavored to develop a high-precision predictive model based on homework record data to analyze students' homework habits through the proposed algorithm in order to help teachers make better improvements to their teaching models. With five homework records and a total of 14 features analyzed, the best CatBoost algorithm achieved a prediction accuracy of 93.34%. In terms of the findings, the degree of procrastination had the biggest influence on students, followed by the degree of reliability and the degree of reliability. The degree of procrastination had the most direct impact on students, and students with high levels of procrastination were generally less motivated to learn. The degree of reliability is used as an indicator to judge students' plagiarism; those who are dishonest are not less motivated to learn but have an easier way to obtain high marks, which is something they are more than happy to see. In summary, it is possible to use work habits to analyze student learning characteristics to adequately predict early academic performance and to provide teachers with decision-support tools.

Student learning is characterized by multiple scenarios and stages. The model proposed by the study needs to be used for different application scenarios and different age groups and to add targeted judgments for judging more types of questions for more refined predictions. The collection of data using multiple forms of homework will be considered to integrate multimodality into the existing model and provide further validation to continuously improve the accuracy of the model.

Author Contributions: Conceptualization, W.W. and Z.Z.; methodology, W.W and Z.Z.; software, Y.S. and W.W.; validation, Y.S. and W.W.; investigation, Z.Z. and Y.L.; data curation, Y.S.; writing—original draft preparation, Y.L.; writing—review and editing, Z.Z. and Y.L.; project administration, Y.L. and Y.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Original data are not publicly available due to ethical restrictions on identifying participants.

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

References

1. Lemay, D.J.; Doleck, T.; Bazelais, P. Transition to online teaching during the COVID-19 pandemic. *Interact. Learn. Environ.* **2021**, *4*, 100130.
2. Martin, F.; Sun, T.; Westine, C. A systematic review of research on online teaching and learning from 2009 to 2018. *Comput. Educ.* **2020**, *159*, 104009. [CrossRef]
3. González, C. The relationship between approaches to teaching, approaches to e-teaching and perceptions of the teaching situation in relation to e-learning among higher education teachers. *Instr. Sci.* **2012**, *40*, 975–998. [CrossRef]
4. Kim, K.; Kim, H.-S.; Shim, J.; Park, J.S. A Study in the Early Prediction of ICT Literacy Ratings Using Sustainability in Data Mining Techniques. *Sustainability* **2021**, *13*, 2141. [CrossRef]
5. Muca, E.; Cavallini, D.; Odore, R.; Baratta, M.; Bergero, D.; Valle, E. Are Veterinary Students Using Technologies and Online Learning Resources for Didactic Training? A Mini-Meta Analysis. *Educ. Sci.* **2022**, *12*, 573. [CrossRef]
6. Michael, P.; Trigwell, K. Relations between perceptions of the teaching environment and approaches to teaching. *Br. J. Educ. Psychol.* **1997**, *67*, 25–35.
7. Liang, K.; Zhang, Y.; He, Y.; Zhou, Y.; Tan, W.; Li, X. Online behavior analysis-based student profile for intelligent E-learning. *J. Electr. Comput. Eng.* **2017**, *2017*, 9720396. [CrossRef]
8. Romero, C.; Ventura, S. Educational data science in massive open online courses. *Wiley Interdiscipl. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1187. [CrossRef]
9. Zamecnik, A.; Kovanović, V.; Joksimović, S.; Liu, L. Exploring non-traditional learner motivations and characteristics in online learning: A learner profile study. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100051. [CrossRef]
10. Tian, A. Comprehensive quality assessment: Changes and implementation of learning assessment in the age of intelligence. *China's Comput. Educ.* **2020**, *6*, 109–113.
11. Avella, J.T.; Kebritchi, M.; Nunn, S.; Kanai, T. Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learn.* **2016**, *20*, 2.
12. Kay, D.; Korn, N.; Oppenheim, C. Legal, Risk and Ethical Aspects of Analytics in Higher Education |. Cetis LLP Publications, Vol 1(No 6). 2012. Available online: https://www.semanticscholar.org/paper/Legal%2C-Risk-and-Ethical-Aspects-of-Analytics-in-Kay-Korn/dce91ba2481a4cb6c849110202de48849e5e3550 (accessed on 30 November 2018).
13. Slade, S.; Prinsloo, P. Learning analytics: Ethical issues and dilemmas. *Am. Behav. Sci.* **2013**, *57*, 1510–1529. [CrossRef]
14. Cui, Y.L.; Zhang, H. Dilemmas, causes and countermeasures of educational artificial intelligence applications. *Mod. Educ. Technol.* **2022**, *32*, 35–42.
15. Yin, B.; Wu, F. Research on modeling of homework habits in intelligent learning system. *Electrochem. Educ. Res.* **2021**, *42*, 61–67. [CrossRef]
16. Yin, B.; Wu, F. Research on the method of modeling learning habits in intelligent learning system. *Electrochem. Educ. Res.* **2020**, *41*, 55–61. [CrossRef]
17. Wu, F.; Yin, B.; Huang, S. Learning Habit Dynamics Research Paradigm and Its Innovative Value. *Mod. Distance Educ. Res.* **2021**, *2019*, 46–52.
18. Yin, B.; Wu, F. Principles and Model Design of Online Intervention of Learning Habits. *e-Educ. Res.* **2019**, *40*, 72–79. [CrossRef]
19. Fogg, B.J. A behavior model for persuasive design. In Proceedings of the Persuasive Technology, Fourth International Conference, PERSUASIVE 2009, Claremont, CA, USA, 26–29 April 2009; pp. 1–7.
20. Flunger, B.; Trautwein, U.; Nagengast, B.; Lüdtke, O.; Niggli, A.; Schnyder, I. A person-centered approach to homework behavior: Students' characteristics predict their homework learning type. *Contemp. Educ. Psychol.* **2017**, *48*, 1–15. [CrossRef]
21. Akram, A.; Fu, C.; Li, Y.; Javed, M.Y.; Lin, R.; Jiang, Y.; Tang, Y. Predicting students' academic procrastination in blended learning course using homework submission data. *IEEE Access* **2019**, *7*, 102487–102498. [CrossRef]
22. Trockel, M.T.; Barnes, M.; Egget, D.L. Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *J. Am. Coll. Health* **2000**, *49*, 125–131. [CrossRef]
23. Xu, J. Models of secondary school students' interest in homework: A multilevel analysis. *Am. Educ. Res. J.* **2008**, *45*, 1180–1205. [CrossRef]
24. King, D.L.; Case, C.J. E-cheating: Incidence and trends among college students. *Issues Inf. Syst.* **2014**, *15*, 20–27.
25. Yu, J.; Li, Y.; Cheng, L.; Lian, S.; Tan, C.; Ding, D.; Liu, Q. Research and practice on the method of plagiarism detection for higher education program code assessments. *J. Univ. Sci. Technol. China* **2020**, *50*, 1048–1057.
26. Yazdanparast, A.; Noori, M.; Khayyat, H.J.; Ashraf azimi, M.; Bolourian, M.; Mirzaei, M. Examining effective factors, inhibitory factors and the most common methods of cheating in students: A systematic review. *Med. Educ. Bull.* **2021**, *2*, 143–152.
27. Ljubovic, V.; Pajic, E. Plagiarism detection in computer programming using feature extraction from ultra-fine-grained repositories. *IEEE Access* **2020**, *8*, 96505–96514. [CrossRef]
28. Yu, H.-Y. The development and reflection of China's school situation analysis research in the past decade or so. *Shanghai Educ. Res.* **2019**, *3*, 60–64. [CrossRef]
29. Gang, L.; Jing, T. Value Implications, Practical Dilemmas and Improvement Paths of Learning Analysis. *Teach. Manag.* **2020**, *27*, 18–21.
30. Zhang, J.; Ma, D.; Yan, Q. Analysis of nursing research course learning in Chinese university catechism platform. *Chin. Nurs. Educ.* **2021**, *18*, 623–626.

31. Motz, B.A.; Mallon, M.; Quick, J.D. Automated Educative Nudges to Reduce Missed Assessment in College. *IEEE Trans. Learn. Technol.* **2021**, *14*, 189–200. [CrossRef]

32. Tamara, M.; Nishen, A.K.; Dickhäuser, O. Students' perception of teachers' reference norm orientation and cheating in the classroom. *Front. Psychol.* **2021**, *12*, 614199.

33. Ullah, F.; Wang, J.; Farhan, M.; Jabbar, S.; Wu, Z.; Khalid, S. Plagiarism detection in students' programming assignments based on semantics: Multimedia e-learning based smart assessment methodology. *Multimed. Tools Appl.* **2020**, *79*, 8581–8598. [CrossRef]

34. Mohammad, S.; Gholampour, S. Cheating on exams: Investigating reasons, attitudes, and the role of demographic variables. *Sage Open* **2021**, *11*, 21582440211004156.

35. Melih, D.; Ferdi, B. University students' views on the effectiveness of learning through homework. *Int. Online J. Educ. Sci.* **2021**, *13*, 689–704.

36. Alexander, E.S.; Onwuegbuzie, A.J. Academic procrastination and the role of hope as a coping strategy. *Personal. Individ. Differ.* **2007**, *42*, 1301–1310. [CrossRef]

37. Murat, B.; Duru, E.; Bulus, M. Analysis of the relation between academic procrastination, academic rational/irrational beliefs, time preferences to study for exams, and academic achievement: A structural model. *Eur. J. Psychol. Educ.* **2013**, *28*, 825–839.

38. Marcílio, W.E.; Eler, D.M. From explanations to feature selection: Assessing SHAP values as feature selection mechanism. In Proceedings of the 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Porto de Galinhas, Brazil, 7–10 November 2020.

39. Son, J.-W.; Noh, T.-G.; Song, H.-J.; Park, S.-B. An application for plagiarized source code detection based on a parse tree kernel. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1911–1918. [CrossRef]

40. Klassen, R.M.; Krawchuk, L.L.; Rajani, S. Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemp. Educ. Psychol.* **2008**, *33*, 915–931. [CrossRef]

41. Tohara, A.J.T. Exploring digital literacy strategies for students with special educational needs in the digital age. *Turk. J. Comput. Math. Educ.* **2021**, *12*, 3345–3358. [CrossRef]

42. Jeong, H.-Y.; Choi, C.-R.; Song, Y.-J. Personalized learning course planner with E-learning DSS using user profile. *Expert Syst. Appl.* **2012**, *39*, 2567–2577. [CrossRef]

43. Kostolányová, K.; Šarmanová, J. Use of adaptive study material in education in E-learning environment. *Electron. J. e-Learn.* **2014**, *12*, 172–182.

44. Magdin, M.; Turčáni, M. Personalization of student in course management systems on the basis using method of data mining. *Turk. Online J. Educ. Technol.* **2015**, *14*, 58–67.

45. Wang, K.H.; Wang, T.; Wang, W.; Huang, S.C. Learning styles and formative assessment strategy: Enhancing student achievement in Web-based learning. *J. Comput. Assist. Learn.* **2006**, *22*, 207–217. [CrossRef]

46. Hsu, M.-H. A personalized English learning recommender system for ESL students. *Expert Syst. Appl.* **2008**, *34*, 683–688. [CrossRef]