



**HAL**  
open science

# Predicting Student Performance in Online Learning Using a Highly Efficient Gradient Boosting Decision Tree

Chang Wang, Liang Chang, Tiejuan Liu

► **To cite this version:**

Chang Wang, Liang Chang, Tiejuan Liu. Predicting Student Performance in Online Learning Using a Highly Efficient Gradient Boosting Decision Tree. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.508-521, 10.1007/978-3-031-03948-5\_41 . hal-04178715

**HAL Id: hal-04178715**

**<https://inria.hal.science/hal-04178715v1>**

Submitted on 8 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Predicting Student Performance In Online Learning Using A Highly Efficient Gradient Boosting Decision Tree

Chang Wang<sup>2</sup>, Liang Chang<sup>1,2</sup>, Tiejuan Liu<sup>2,3</sup>

<sup>1</sup> Guangxi Key Laboratory of Trusted Software, Guilin 541004, China

<sup>2</sup> School of Computer Science and Information Security,  
Guilin University of Electronic Technology, Guilin 541004, China

<sup>3</sup> School of Information and Communication,  
Guilin University of Electronic Technology, Guilin, 514000 China  
lty205@guet.edu.cn

**Abstract.** Online learning has become a popular way of learning due to the rapid development of informatization of education, The rise of online learning platforms has provided great convenience to students' learning, breaking the limitations of time and space, enabling students to study various courses anytime, anywhere. However, due to the lack of an effective supervision mechanism, online learning undermined by low quality tutoring and a substantial dropout rate of students. In this study, we used an efficient ensemble learning algorithm, LightGBM, to develop a student performance prediction model. Basing on the interaction behavior data of students' online learning, the prediction model can predict whether a student will be able to pass the course. As a result of identifying at-risk students, teachers can provide targeted interventions to improve the learning performance of these students. During the experiments, we compared with ten classical machine learning algorithms on a public dataset. It reflected that LightGBM outperformed in predicting student performance. The study also analyzed the impact of different interaction behaviors on students' online learning performance, which will help teachers and educational researchers to better understand the relationship between students' online learning behavior and students' learning performance. This will help teachers guide students in online learning and optimize the design of online learning courses.

**Keywords:** LightGBM, Student Performance Prediction, Learning Analytics.

## 1 Introduction

In recent years, in conjunction with the rapid development of Internet technology, a large number of online learning systems have emerged. These include Virtual Learning Environments (VLEs), Massive Open Online Courses (MOOCs), etcetera. This innovation enables there to be intellectual freedom and also allows students to take courses without going to an educational institution. However, online learning faces a multitude of challenges. The number of students studying online far exceeds that of

those taking traditional courses, and it is difficult for teachers to track student learning, which makes many under-performing students drop out of courses due to lack of timely guidance[1]. Fortunately, these online learning platforms widely store a large amount of student learning data, which benefits educational research immensely[2]. As a result of analyzing the learning data of students during the online learning process, it is possible to predict students' learning performance, help teachers identify at-risk students, and provide targeted additional help and interventions to improve students' academic performance.

Predicting student performance as an effective measure to identify at-risk students has been recognized as an important research problem in online learning [4]. Accurate predictions of student performance will help provide students with personalized guidance to help them pass the course, all the while, obtaining effective predictions requires powerful predictive models [5]. Therefore, many researchers have used various machine learning methods to develop the best predictive models of student performance. Among these methods, ensemble learning can often achieve better prediction results. As an algorithm of ensemble learning, random forest has been widely used in student performance prediction research and achieved excellent prediction results [6].

The student performance prediction task in the online learning environment faces new challenges which are different from the student performance prediction task in the traditional course learning environment. Due to the convenience and freedom of learning, online platforms attract a large number of students, which makes online platforms record a large amount of data related to student learning. However, due to the lack of constraints on student learning, this data often exhibits high-dimensional and sparse characteristics. It is difficult for traditional ensemble learning algorithm to analyze quickly and efficiently. LightGBM not only can handle large datasets, but also has high predictive ability, as an efficient ensemble algorithm. The algorithm can also utilize the sparsity of features to perform lossless merging of features, which is suitable for processing high-dimensional and sparse data. Therefore, we use the LightGBM algorithm to build a student performance prediction model, which predicts students' performance by analyzing their interaction data with the platform during their online learning process. In addition to that, we analyze the importance of different features in prediction and the correlation of features to explore the relationship between students' different interaction behaviors in the learning process and student performance. Based on these explorations, it can provide effective assistance to teachers to guide students' learning, and provide meaningful reference for the design of online courses.

## **2 Literature review**

In the field of student performance prediction, students' learning performance prediction tasks are generally divided into two categories: classification tasks and regression tasks. The classification task is to predict whether a student will pass a course or predict a student's final grade. The regression task is the prediction of the most likely

score a student will get in the final exam [7]. Traditional machine learning methods are widely used to predict student performance with good results, such as regression analysis [8], decision trees [9, 10], Bayesian [11], support vector machines [12, 13, 14], neural network [15, 16] and other methods. These algorithms are often based on a single model for predicting student performance, which may perform differently for different datasets, making it difficult to identify a generally accepted prediction algorithm [17]. As a result of widespread application of ensemble learning in other fields, more and more researchers apply ensemble learning to student performance prediction. Compared with single model prediction methods, ensemble learning can assemble multiple weak classifiers into a strong classifier, improve the performance of prediction, and can achieve good prediction results on datasets of different scales.

The most commonly used ensemble learning in student performance prediction is random forest, an ensemble learning algorithm developed by Breiman and Adele Cutler, which uses a large number of decision tree models to improve accurate predictions. It does this by reducing the bias and variance of estimate [18]. Its essence is a random combination of freely growing decision trees in a random subspace through Bagging. Trakunphutthirak et al. [19] collected students' online logs for one week during the exam, which included 90 students' online activities during the exam, including 120 Internet access activities and 25 browsing categories. Process logs into two distinct datasets by visit activity and browsing category. On the two datasets, students' learning performance is predicted based on the student's usage time and the student's usage frequency, respectively. The experimental results show that the random forest is the best method for predicting the risk of students' failure compared to decision trees, naive Bayes, logistic regression and neural networks. Hasan et al. [20] used video learning data and data mining techniques to predict whether a student would be able to pass a course. The dataset consists of academic data in SSI, activity data in Moodle, and video clickstream data in eDify for 772 students during an academic year. The commonly adopted classification algorithms in the existing literature are used: random forest, k-nearest neighbor, support vector machine, etc. The results show that random forest is the best amongst the other methods with an accuracy rate of 88.3%.

As a meta-algorithm in ensemble learning, bagging can combine different models to reduce the generalization error and effectively improve the prediction accuracy [21]. Injada et al. [22] proposed a method to systematically build a bagging ensemble learner for multiple segments, which selects appropriate weak classifiers according to the Gini coefficient and the target classification score. Basing on this method, six basic machine learning algorithms were used to combine the optimal bagging ensemble learner to predict the performance of students in different stages of the course in the e-learning environment. As a result of conducting sufficient experiments on different undergraduate datasets from two universities, the established Bagging ensemble learner has higher prediction accuracy compared with the classical classification techniques. In order to predict the performance of students in academic competitions, Yan et al. [23] used support vector machines, random trees and AdaBoost as basic learners to predict students' performance respectively, and used a simple logistic regression algorithm as a voter to integrate the prediction results of different base learn-

ers. On the experimental data set, through the comparison of five evaluation indicators, the model not only outperforms the commonly used single machine learning algorithm, but also outperforms other ensemble learning algorithms.

As another major class of ensemble learning algorithms, Boosting algorithms have also been used in student performance prediction research. Contrast to Bagging, the Boosting algorithm assembles weak classifiers into a strong classifier by reducing bias. Han et al. [24] used a decision tree-based AdaBoost algorithm to predict the final grades of students in compulsory courses. The ensemble algorithm was evaluated using performance data from 123 undergraduate students in 20 compulsory courses over four years. The experimental results show that the algorithm has better prediction accuracy compared with decision tree, neural network, random forest and other models. Ahmed et al. [25] proposed the use of the GBDT algorithm to predict student performance on final exams, comparing with support vector machines, logistic regression, and naive bayes on a dataset containing 450 undergraduate students. The experimental results show that the algorithm has better performance than other algorithms, and the prediction accuracy is 89.1%.

In the existing research on student performance prediction, the prediction effect of ensemble learning is often better than that of single machine learning algorithms. However, when the dataset features are high-dimensional and the number of samples is large, the efficiency and scalability of the existing ensemble learning algorithms are unsatisfactory. As an efficient ensemble learning method, LightGBM can be applied to large-scale datasets with high feature dimensions, and achieves higher accuracy on multiple public datasets. So in this article, we will use LightGBM to predict student performance. Although Vultureanu-Albiși et al. [26] used different ensemble learning algorithms based on tree models to predict student performance, they found that decision trees outperformed other ensemble learning algorithms including LightGBM. However, the two datasets used in this study are both small, with only a few hundred students, and cannot well demonstrate the predictive performance of the ensemble learning algorithm in the online learning environment. We used LightGBM to predict student online learning performance with a dataset containing several thousand students. Furthermore, it will be compared with a variety of classic machine learning algorithms to evaluate the performance of the algorithm in the task of predicting student performance.

### **3 Methodology**

LightGBM is an efficient ensemble learning algorithm proposed by Guolin et al. [27]. The purpose is to solve the problem that the efficiency and scalability of the existing GBDT algorithm are still unsatisfactory when the feature dimension is high and the amount of data is large. From the perspective of reducing the number of data instances and features during model training, two new techniques in the LightGBM, Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), are proposed to improve GBDT.

### 3.1 Gradient-based One-Side Sampling

In GBDT, the gradient information of each data instance can provide valuable information for data sampling. Instances with larger gradients (that is, instances with insufficient training) contribute more to the information gain. However, reducing the data training size by discarding data instances with small gradients will change the data distribution and affect the accuracy of the trained model. The GOSS algorithm can randomly remove instances with small gradients without changing the original data distribution immensely. When the algorithm determines the sample set for a new round of training, it will sort the samples by the absolute value of the gradient according to the current training result. All instances with large gradients are used as selected samples, and a certain proportion of instances with small gradients are randomly selected as the ideal ones, thus selected. In order not to change the original distribution by a great margin, the ratio of the number of selected small gradient samples to all small gradient samples needs to be used as the weight coefficient of the selected small gradient samples.

---

**Algorithm 1** Gradient-based One-Side Sampling

---

**Input:**  $I$ : training data,  $d$ : iterations  
**Input:**  $a$ : Sampling ratio of large gradient data  
**Input:**  $b$ : Sampling ratio of small gradient data  
**Input:**  $loss$ : loss function,  $L$ : weak learner  
 $models \leftarrow \{\}, fact \leftarrow \frac{1-a}{b}$   
 $topN \leftarrow a \times len(I), randN \leftarrow b \times len(I)$   
**for**  $i = 1$  **to**  $d$  **do**  
     $preds \leftarrow models.predict(I)$   
     $g \leftarrow loss(I, preds), w \leftarrow \{1, 1, \dots\}$   
     $sorted \leftarrow GetSortedIndices(abs(g))$   
     $topSet \leftarrow sorted[1:topN]$   
     $randSet \leftarrow RandomPick(sorted[topN:len(I)], randN)$   
     $usedSet \leftarrow topSet + randSet$   
     $w[randSet] \times = fact$   $\triangleright$  Assign weight fact to the small gradient data  
     $newmodel \leftarrow L(I[usedSet], -g[usedSet], w[usedSet])$   
     $models.append(newModel)$

---

### 3.2 Exclusive Feature Bundling

Since it is evident in many cases that the data is sparse with high feature dimension, EFB can reduce the number of features without loss, and can speed up the training speed of GBDT under the premise of ensuring accuracy. EFB consists of two parts: Identification of features that should be bundled together and construction of the bundle.

In the first part, it is not only necessary to find suitable bundling features, but the number of bundles should be minimal, therefore a reasonably greedy algorithm is ideal and used. Since there are more non-zero values of the features, there may be more conflict, therefore the greedy algorithm will sort in descending order according to the number of non-zero values of the features, and then check each feature according to the sorted sequence, and assign it to a smaller conflicting bundle or create a new one.

---

### Algorithm 2 Greedy Bundling

---

**Input:**  $F$ : features,  $K$ : max conflict count  
 Non zero statistical set  $S$   
 $searchOrder \leftarrow S.sortByNumberOfNonZero()$   
 $bundles \leftarrow \{\}, bundlesConflict \leftarrow \{\}$   
**for**  $i$  **in**  $searchOrder$  **do**  
    $needNew \leftarrow True$   
   **for**  $j = 1$  **to**  $len(bundles)$  **do**  
      $cnt \leftarrow conflictCnt(bundles[j], F[i])$   
     **if**  $cnt + bundlesConflict[i] \leq k$  **then**  
        $bundles[j].add(F[i], needNew \leftarrow False)$   
       **break**  
   **if**  $needNew$  **then**  
     **Add**  $F[i]$  as a new bundle to  $bundles$   
**Output:**  $bundles$

---

In the second part, in order to reduce the training complexity and be able to identify the original features from the feature bundles, a feature bundle is constructed by increasing the offset of the original value of the feature.

---

### Algorithm 3 Merge Exclusive Features

---

**Input:**  $numData$ : number of data  
**Input:**  $F$ : One bundle of exclusive features  
 $binRanges \leftarrow \{0\}, totalBin \leftarrow 0$   
**for**  $f$  **in**  $F$  **do**  
    $totalBin += f.numBin$   
    $binRanges.append(totalBin)$   
 $newBin \leftarrow new\ Bin(numData)$   
**for**  $i = 1$  **to**  $numData$  **do**  
    $newBin[i] \leftarrow 0$   
   **for**  $j = 1$  **to**  $len(F)$  **do**  
     **if**  $F[j].bin[i] \neq 0$  **then**  
        $newBin[i] \leftarrow F[j].bin[i] + binRanges[j]$   
**Output:**  $newBin, binRanges$

---



Compared with other ensemble learning algorithms, when the feature dimension is high and the amount of data is large, LightGBM is more suitable and often achieves better prediction results. So, we use LightGBM to predict whether students would be able to pass the online course. We use ten classic machine learning algorithms to evaluate the predictive performance of LightGBM on student performance tasks, including Logistic Regression(LR), K-Nearest Neighbors(KNN), Decision Trees(DT), Naive Bayes(NB), Support Vector Machines(SVM), Multi-layered Perceptron(DNN), Random Forests(RF), Adaboost, GBDT, XGBoost. Among these algorithms, XGBoost and LightGBM are implemented based on the Sklearn interface in their respective python packages, while other algorithms are implemented using the Sklearn toolkit.

## 4 Experiments

### 4.1 Experimental setup

#### Dataset

The Open University Learning Analytics Dataset (OULAD) is a dataset containing aggregated clickstream data of student interactions with the learning environment in a virtual learning environment (VLE) [28]. This dataset is often used to evaluate the predictive performance of student performance prediction models [29, 30]. In order to avoid the influence of course type and effectively evaluate the performance of the model, we use the data of the FFF courses with the largest number of students and the most frequent interaction behaviors for experimental evaluation.

The course's original dataset contains more than 4 million interactions generated by 7,762 students. Among them, some students participated in courses multiple times, some students also participated in other courses, and some students' interaction data with the platform was not recorded. Therefore, we deleted the data of these students, and finally generated an experimental data set containing 6455 students and their interaction data with the platform. We aggregated the interaction data of students during the learning process by interaction type to generate 18 features for prediction.

#### Evaluation Metrics

In this experiment, we use five categories of metrics, Accuracy, Precision, Recall, F1-Score, and AUC to evaluate the performance of the student performance prediction method.

**Accuracy:** The percentage of students who predict the correct performance of the total students. Although the accurate rate can judge the correct rate of the prediction, it cannot be used as a good measure when the sample is unbalanced.

**Precision:** The percentage of students who actually passed the course among all the students who were predicted to pass the course.

**Recall:** The percentage of students who passed the course that were correctly predicted in the real world.

F1-Score: The harmonic mean of precision and recall. The calculated value is between 0 and 1. The larger the F1 value, the better the classification quality.

AUC: The area under the ROC curve. Generally, the AUC is between 0.5 and 1. The larger the value, the better the effect. It is often used to measure the pros and cons of learners.

#### 4.2 Comparison of LightGBM and the single machine learning algorithms

To achieve this research goal, we used six single machine learning classification algorithms commonly used in student performance prediction research to separately predict whether students will pass the course, and the results are shown in Table 1.

**Table 1.** Comparison with the single machine learning algorithms

	Accuracy	Precision	Recall	F1-Score	AUC
NB	0.783	0.906	0.671	0.771	0.794
KNN	0.869	0.914	0.837	0.874	0.871
DT	0.904	0.922	0.899	0.91	0.904
LR	0.907	<b>0.924</b>	0.903	0.914	0.907
SVM	0.931	0.901	<b>0.979</b>	0.939	0.926
DNN	0.937	0.916	0.973	0.944	0.934
LightGBM	<b>0.941</b>	<b>0.924</b>	0.972	<b>0.947</b>	<b>0.938</b>

From Table 2, we can find that in the single algorithm model, the NB algorithm achieves the worst results. The NB algorithm assumes that each feature in the sample is independent of each other, and in the learning process, there are often correlations between different types of interaction behaviors of students, which are not completely independent. DNN algorithm has strong fitting ability, can fit any complex nonlinear relationship in theory, and can fully learn the relationship between features, so it achieves better prediction results. As an ensemble learning method, LightGBM integrates multiple decision tree classifiers, each of which is trained on the basis of the residual of the previous round of classifiers. During the training process, the bias is continuously reduced to improve the accuracy of the classifier. Compared with single algorithms, LightGBM achieves the best prediction performance.

#### 4.3 Comparison of LightGBM with other classical ensemble learning methods

In this experimental comparison, we use four commonly used ensemble learning algorithms to predict whether students can pass the course. These ensemble algorithms are all implemented based on decision trees. Among them, RF belongs to the extension of Bagging in the ensemble learning algorithm, and performs random feature selection on the basis of Bagging with decision tree as the base learner. AdaBoost, GBDT, and XGBoost all belong to Boost-type ensemble learning. The difference is that AdaBoost will process the samples from the previous training, increase the weight of wrongly classified samples and reduce the weight of correctly classified samples, and generate

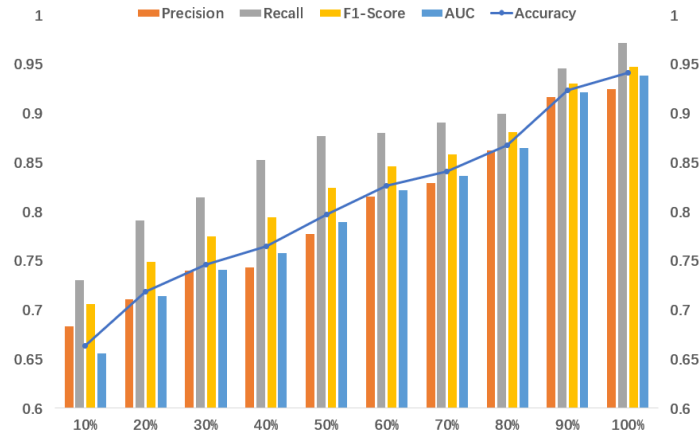
new training samples for the next classifier. Both GBDT and XGBoost train new classifiers through the residuals of the previous round of classifiers. XGBoost adds the complexity of the tree model to the regular term, so the generalization performance is better than GBDT. The prediction effects of these ensemble learning methods are shown in Table 2.

**Table 2.** Comparison with other ensemble learning algorithms

	Accuracy	Precision	Recall	F1-Score	AUC
RF	0.934	0.904	<b>0.983</b>	0.942	0.93
AdaBoost	0.929	0.909	0.966	0.936	0.925
GBDT	0.9356	0.912	0.976	0.943	0.932
XGBoost	0.937	0.919	0.97	0.944	0.934
Lightgbm	<b>0.941</b>	<b>0.924</b>	0.972	<b>0.947</b>	<b>0.938</b>

#### 4.4 Predictive performance of LightGBM at different stages of the course

In order to better study the prediction performance of LightGBM and achieve the purpose of early prediction, we divide the courses into different course stages by time, such as 10%, 20%, ..., 100%. Then, the prediction model based on LightGBM is used to predict the performance of students under different course stages respectively.



**Fig. 1.** Predictive performance of LightBDM at different course stages

From Figure 1, we can easily find that as the course progresses, the accuracy of student predictions will gradually improve. As there is continuous development of the course, the data of students interacting with the online platform continues to increase, which helps the model to make more accurate predictions. Among them, when the course reaches 80% and above, we can make more accurate predictions for students.

#### 4.5 Temporal performance evaluation of LightGBM

As opposed to the student performance prediction task in the traditional learning environment, the student performance prediction task in the online learning environment has the characteristics of big data, sparse and high dimensional features, which makes the model training time longer. Therefore, it is equally important to reduce the training speed of the model when ensuring the predictive performance of the model. We calculated the time of different models on the student performance prediction task, and the prediction performance of these models was similar to LightGBM. From Figure 2, we find that LightGBM takes the shortest time on the student performance prediction task.

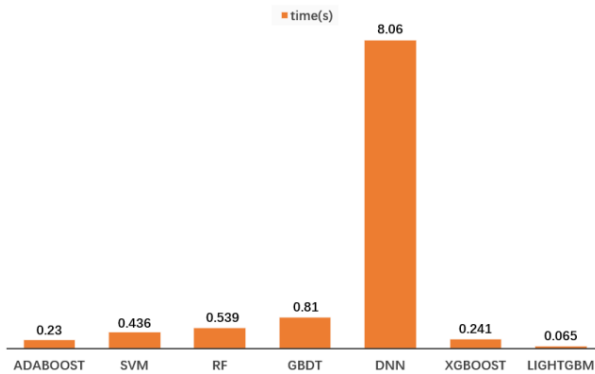
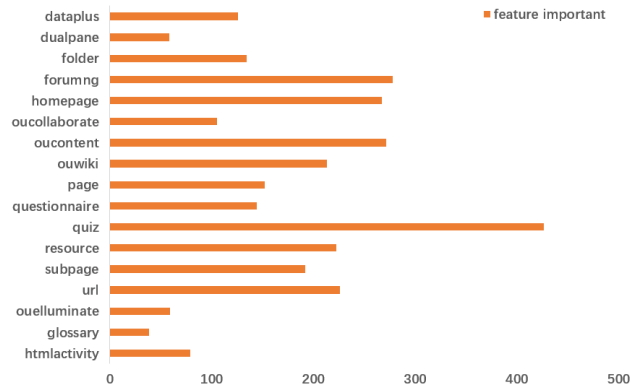


Fig. 2. Temporal performance of different prediction algorithms

#### 4.6 Interaction types importance analysis

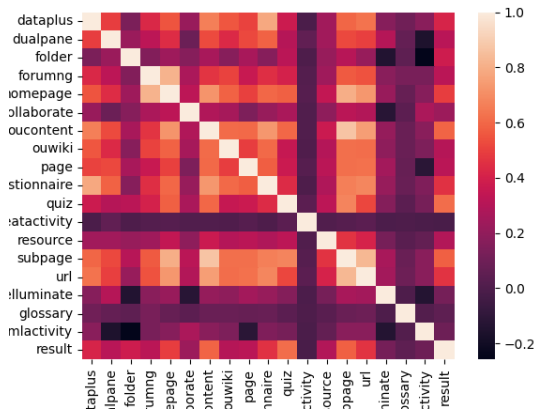
While choosing a model that accurately predicts student performance is important, so is the interpretability of the model. Using LightGBM as a predictive method for student performance, it is possible to analyze the importance of different features to the prediction target. In the student performance prediction task, the analysis of feature importance not only provides guidance for teachers to help at-risk students, but also helps to optimize the design of online courses.



**Fig. 3.** Importance of features when predicting student performance

From Figure 3, it can be seen that quiz is crucial for students to pass the course, because a lot of quiz interaction means that students will have a lot of quiz practice, which helps the students to pass the course. Forumng, homepage, and oucontent also plays a very important role. Among them, the visit to forumng means that students will conduct a lot of communication and discussion through the forum, the visit to homepage indicates that the students have a significant interest in the study of the course, the visit to the oucontent indicates that the students have the habit of completing the homework conscientiously.

In order to verify the analysis of the importance of the features by the LightGBM algorithm, we also use the Pearson correlation coefficient to analyze the correlation between different features and between the features and the final result.



**Fig. 4.** A heatmap showing correlations between features and the final result.

On the heatmap of the correlation analysis, it increases as the color becomes lighter. We found that quiz and results were the lightest on the heatmap, reflecting that they had the highest correlation. Oucontent, homepage and results are also lighter in color

on the heatmap, which means they are strongly correlated. However, through observation, we also found that there are differences with LightGBM analysis. From the heatmap, we can see that subpage, link, and file are more strongly correlated with results than forumng, but the importance of analysis in LightGBM is lower than forumng. By observing the correlations between different features, we found that forumng is more strongly correlated with other features than subpage, link, and file. Among them, file even has negative correlations with other features. Therefore LightGBM will give a higher importance to forumng. Similarly, we also verified from here that when using the interaction between students and the platform to predict students' online learning performance, the prediction results of Naive Bayes are the worst, as compared to that of neural networks, which offered much better results. Because there are complex correlations between interactions.

## 5 Conclusions

In this paper, we provide a new perspective on student performance prediction in an online learning environment. In the online learning environment, the datasets used for student performance prediction have scalability problems, sparse and high dimensional features, as well as the time performance requirements for prediction models in practical applications. It is proposed to build a model using the LightGBM algorithm to predict students' ability to pass online learning courses, which is more suitable for predicting student performance in the online learning environment. Finally, six classic single machine learning algorithms and four other commonly used ensemble learning algorithms are compared to predict student performance. Experiments show that the prediction model using LightGBM algorithm not only has high accuracy, but also offers better time performance.

In addition, in order to study the impact of students' online learning interaction behavior on students' learning performance, the importance and correlation of different features were analyzed. We have found that taking more quizzes is essential for students to pass the course. At the same time, forum communication, maintaining interest in the course, and completing assignments can help students through the course. By analyzing the impact of different interaction behaviors on students' learning performance, it can help teachers to effectively guide students' learning, and help online learning courses to be reasonably and effectively designed.

## Acknowledgements

This work was supported by the Natural Science Foundation of China (Nos.U1811264, 62066010), the Natural Science Foundation of Guangxi Province (No.2020GXNSFAA159055), Innovation Project of Guang Xi Graduate Education(No.YCBZ2021072), Guangxi Key Laboratory of Trusted Software (No.KX202058)

## References

1. Chiu, Yu-Chen, et al. : Predicting Student Performance in MOOCs Using Learning Activity Data. *Journal of Information Science and Engineering* 34(5),1223-1235 (2018).
2. Hernández-Blanco, Antonio, et al.: A systematic review of deep learning approaches to educational data mining. *Complexity* (2019).
3. Na, Kew Si, and Zaidatun Tasir.: Identifying at-risk students in online learning by analysing learning behaviour: A systematic review. 2017 IEEE Conference on Big Data and Analytics (ICBDA). IEEE, 118-123(2017).
4. Macfadyen, Leah P., and Shane Dawson.: Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & education* 54 (2): 588-599(2010).
5. Asiah M, Zulkarnaen K N, Safaai D, et al.: A review on predictive modeling technique for student academic performance monitoring. *MATEC Web of Conferences*. EDP Sciences, 255: 03004. (2019).
6. Namoun, Abdallah, and Abdullah Alshantqi.: Predicting student performance using data mining and learning analytics techniques: a systematic literature review. *Applied Sciences* 11(1), 237(2021).
7. Tomasevic, Nikola, Nikola Gvozdenovic, and Sanja Vranes.: An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education* 143: 103676(2020).
8. Zhang W, Huang X, Wang S, et al.: Student performance prediction via online learning behavior analytics. 2017 International Symposium on Educational Technology (ISET). IEEE, 153-157(2017).
9. Liu W, Wu J, Gao X, et al.: An early warning model of student achievement based on decision trees algorithm. 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE). IEEE, 517-222(2017).
10. Wang G H, Zhang J, Fu G S.: Predicting student behaviors and performance in online learning using decision tree. 2018 Seventh International Conference of Educational Innovation through Technology (EITT). IEEE, 214-219(2018).
11. Quan W, Zhou Q, Zhong Y, et al.: Predicting at-risk students using campus meal consumption records. *The International journal of engineering education*, 35(2): 563-571(2019).
12. Burman I, Som S.: Predicting students academic performance using support vector machine. 2019 Amity International Conference on Artificial Intelligence (AICAI). IEEE, 756-759.(2019)
13. Al Mayahi K, Al-Bahri M.: Machine Learning Based Predicting Student Academic Success. 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE, 264-268(2020).
14. Yang Y, Hooshyar D, Pedaste M, et al.: Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, 24(24): 18777-18793(2020).
15. Widyahastuti F, Tjhin V U.: Predicting students performance in final examination using linear regression and multilayer perceptron. 2017 10th International Conference on Human System Interactions (HSI). IEEE, 188-192(2017).
16. SANDOVAL I P, Naranjo D, Gilar R, et al.: Neural network model for predicting student failure in the academic leveling course of Escuela Politécnica Nacional. *Frontiers in Psychology*, 11: 3383(2020).

17. Karalar H, Kapucu C, Gürüler H.: Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education*, 18(1): 1-18(2021).
18. Breiman L.: Random forests. *Machine learning*, 45(1): 5-32(2001).
19. Trakunphutthirak R, Cheung Y, Lee V C S.: Detecting student at risk of failure: a case study of conceptualizing mining from internet access log files. 2018 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 365-371(2018).
20. Hasan R, Palaniappan S, Mahmood S, et al.: Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, 10(11): 3894(2020).
21. Breiman L.: Bagging predictors[J]. *Machine learning* , 24(2): 123-140(1996).
22. Injadat M N, Moubayed A, Nassif A B, et al.: Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*, 50(12): 4506-4528(2020).
23. Yan L, Liu Y.: An ensemble prediction model for potential student recommendation using machine learning[J]. *Symmetry*, 12(5): 728(2020).
24. Han M, Tong M, Chen M, et al.: Application of Ensemble Algorithm in Students' Performance Prediction. 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). IEEE, 735-740(2017).
25. Ahmed D M, Abdulazeez A M, Zeebaree D Q, et al.: Predicting University's Students Performance Based on Machine Learning Techniques. 2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS). IEEE, 276-281(2021).
26. Vultureanu-Albiși A, Bădică C.: Improving Students' Performance by Interpretable Explanations using Ensemble Tree-Based Approaches. 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI). IEEE, 215-220(2021).
27. Ke G, Meng Q, Finley T, et al.: Lightgbm: A highly efficient gradient boosting decision tree[J]. *Advances in neural information processing systems*, 30: 3146-3154(2017).
28. Kuzilek J, Hlosta M, Zdrahal Z.: Open university learning analytics dataset. *Scientific data*, 4(1): 1-8(2017).
29. Waheed H, Hassan S U, Aljohani N R, et al.: Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human behavior*, 104: 106189(2020).
30. Rivas A, Gonzalez-Briones A, Hernandez G, et al.: Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, 423: 713-720(2021).