

MOOC Performance Prediction and Online Design Instructional Suggestions Based on LightGBM

Yimin Ren¹, Jun Wang^{1,2(\boxtimes)}, Jia Hao^{1,2}, Jianhou Gan^{1,2}, and Ken Chen¹

¹ Key Laboratory of Education Informatization for Nationalities, Ministry of Education, Yunnan Normal University, Kunming 650500, China dongling1988@126.com

² Yunnan Key Laboratory of Smart Education, Yunnan Normal University, Kunming 650500,

China

Abstract. Recently, the research on the teaching effectiveness of online teaching has gradually become the focus of people's attention. As a well-known learning platform, MOOC has also become the main front for many learners to conduct online learning. However, some students are not clear about their own learning situation during the learning process, so that they can't get a qualifying grade in a MOOC course. Thus, In order to make the teachers and learners to anticipate the learning performance and then check the gaps as early as possible, we propose a MOOC performance prediction model based on the algorithm of LightGBM, and then compare its results with the others state-of art machine learning algorithms. Experimental results show that our proposed method outperform than the others. Research also leverages LightGBM's interpretability, Analyzed the external environment and existing cognitive structure, two factors that affect online learning. And this is the basis for suggestions on online instructional design.

Keywords: Performance prediction · LightGBM · Online instructional design

1 Introduction

Since the 21st century, blending learning has created a new wave of learning internationally, learners are no longer limited to the courses and time offered by the school, and can learn according to their own circumstances. An epidemic in 2020 has brought online education into the spotlight at an alarming rate, followed by researchers' enthusiasm for online education researches. At present, there are many types of online education platforms, but the quality is uneven. One of the most widely used courses in universities is the Massive Open Online Course (MOOC), which is set up by top international universities and provides free courses from various schools on-line. Since its inception in 2013, MOOCs in China have gradually established a unique development model. At present, the number and scale of application of MOOCs in China ranks first in the world. If we can explore the student management method of MOOC and then apply it to domestic MOOCs. Maybe it can help Chinese MOOCs develop better.

[©] The Author(s), under exclusive license to Springer Nature Switzerland AG 2023 Y. Xu et al. (Eds.): ML4CS 2022, LNCS 13657, pp. 507–519, 2023. https://doi.org/10.1007/978-3-031-20102-8_39

Because today's MOOCs are used by learners from all fields of life, everyone learns for different purposes and in different ways. So, many learners may fail to achieve excellent or qualified grades, or even dropout. There are many reasons for this, such as objective reasons like the inability to squeeze out time and the network environment does not allow. As well as subjective reasons like lack of self-control, inability to interact with teachers and give up because of lack of attention. If the MOOC designers can anticipate learners' final grade, they can send some warning based on the prediction results, and then pay attention to check the gaps as soon as possible to avoid failing the course. Teachers can also get certain early warnings and adjust the teaching design as soon as possible to help learners better acquire knowledge.

Although some researchers $[1-3]$ $[1-3]$ are currently conducting related research on MOOC student performance prediction. However, there are a large number of MOOC learners, they generate a large amount of learning behavior data and feature dimensions. Therefore, researchers need to spend a lot of time and storage costs when building a performance prediction model. On the other hand, in the process of online learning, in addition to learners, there are also important members, namely teaching organizers. As one of the components of human resources in learning resources, teachers are ignored by some researchers. After some researchers predict MOOC scores, they only feedback the results to students, allowing students to explore ways to improve their academic performance. They don't understand some machine learning algorithms that provide a ranking of features in the process of building a performance prediction model. It is the ranking of the importance of the features to the prediction results. Corresponding analysis of this ranking can provide new ideas for the teaching design of online education. The fundamental purpose of instructional design is to create a variety of effective teaching systems through the systematic arrangement of teaching processes and teaching resources to promote learners' learning. Unlike traditional teaching, the learners of online teaching may have different a prior knowledge and are in different external environments, but the fundamental purpose of the instructional design is to facilitate learners' learning. Therefore, it is worthwhile to pay attention to how to improve traditional instructional design according to the characteristics of online teaching and learning.

In terms of prediction, The Gradient Boosting Decision Tree (GBDT) in machine learning already has high accuracy and interpretability, and has advanced performance in many tasks. However, when faced with massive data, its shortcomings are also obvious. Because GBDT needs to traverse all instances of each feature to calculate the information gain and find the optimal split point, it is very time-consuming. To make up for this shortcoming of GBDT, LightGBM was proposed. LightGBM proposes two new algorithms from reducing the number of samples and reducing the feature dimension. Experiments show that LightGBM can reduce the prediction time and still have high accuracy.

According to this situation, first of all, this study selects the Open University Learning Analysis Dataset (OULAD), which has a huge amount of data, to build an online performance prediction model based on the LightGBM algorithm. And use Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Adabo-ost (Ada) four machine learning algorithms to build a performance prediction model and compare it with the LightGBM algorithm. The experiment found that the prediction error of the MOOC performance prediction model based on the LightGBM algorithm is smaller. Then, the second work done in this study is, Using the interpretability of LightGBM, the importance ranking table of the influence of features on the prediction results is obtained. According to this ranking, two factors that can provide reference for online teaching design are obtained, namely, the influence of the external environment and the influence of the learner's original cognitive structure. Finally, two online teaching design suggestions are given.

2 Related Research

2.1 MOOC Performance Prediction

The research on performance prediction has become one of the research hotspots in the field of educational technology, but in the process of research, many scholars have to face the problem of excessive time cost caused by the large amount of data. Some researchers avoid this by selecting a small number of subjects. Hasan [\[4\]](#page-11-2) et al. Collected course data from only 22 students and used three types of data characteristics, namely, GPA, pre-course grades, and online module quizzes, to predict the final grade level; Zhao [\[5\]](#page-11-3) et al. selected 38 students of education technology in Northeast Normal University as the research object, and used 13 variables of learning activity as the characteristics of the dataset, designed and constructed an online learning intervention model based on data mining algorithm and learning analysis technology. In addition to the large amount of data generated by learners, another feature of online learning is that learners have many feature dimensions. For example, Li Shuang et al. [\[6\]](#page-12-0). Selected 150 students from four courses at the National Open University as the study population and collected 26 learning variables about their four courses as features to construct the assessment model. Therefore, the author believes that the achievement prediction for MOOC cannot get around the difficult problem of wide data volume and many data set features. Therefore, finding algorithms that can reduce time and storage costs can effectively solves this challenge.

2.2 Analysis of Relevant Factors Affecting the Prediction Results of Online Performance

Soon after the rise of online learning, some researchers focused on the analysis of factors affecting online learning. However, at that time, due to technical limitations, researchers could only use statistical and data mining methods to simply analyze students' learning behaviors to derive relevant factors affecting online learning. For example, Wei et.al [\[7\]](#page-12-1) studied three types of behavioral data generated by 9369 new students studying the online course as an example, and used data mining methods to derive the characteristics of several online learning behaviors and their influencing factors. However, today, the analysis of factors affecting online learning can be incorporated into the study of online learning performance prediction. Because some of the machine learning algorithms are interpretable in the process of building performance prediction models, researchers can use this interpretable component to give suggestions for instructional design enhancements for online learning after the models are built. For example, Qing Wu et al. [\[8\]](#page-12-2). Chose Bayesian networks for student performance prediction and analyzed the model, and then gave suggestions that learners should be more motivated in the learning process. Although Bayesian algorithm is also interpretable, it requires all data to be discretized, so using Bayesian algorithm for grade prediction is more suitable for scenarios where the grade result is a classification problem, such as predicting a pass or fail result.

2.3 Lack of Recommendations for Instructional Organizers at the End of Online Performance Prediction

In order for the study to have a landing point, i.e., a specific application in an educational scenario, the researcher typically uses the prediction results as a reminder to the teachers and learners at the end of the prediction, acting as an early warning. For example, Xu et al. [\[9\]](#page-12-3) predicted student performance and provided early warning of student performance based on a Heterogeneous information network. Some scholars have also made performance predictions early in students' learning so that learners can gain an understanding of their learning early and adjust their learning early [\[10–](#page-12-4)[12\]](#page-12-5). However, simple warnings only provides the limited positive guidance for teachers to change their instructional design.

However, there are still many shortcomings in the current study, including a small number of research subjects so it cannot provide guidance for MOOC courses with a large amount of data, a lack of analysis of explainable factors in the regression prediction process, and some of the studies only serve as an early warning and cannot suggest improvements for teachers and students. Starting from these problems, this study selects the LightGBM algorithm to construct a performance prediction model for a MOOC dataset with a large amount of data: OULAD. Because the online learning dataset has a large amount of data and many feature dimensions, the LightGBM algorithm can take advantage of its own advantages to reduce the amount of data and feature dimensions. This not only ensures the accuracy, but also saves the time and memory overhead of the experiment. Moreover, we can also use the interpretability of the LightGBM algorithm to analyze the prediction results. And propose instructional design improvements to the instructional organizer.

3 Construction of MOOC Learning Performance Prediction Model Based on LightGBM

LightGBM is an algorithm improved on GBDT. GBDT uses a decision tree as a base learner to improve the performance of the model through the idea of integration. The specific workflow of GBDT is shown in Fig. [1](#page-4-0) below. It will input the training data into the weak learner to calculate the gradient through the loss function, and then input the result to the next weak learner. In this process, the weights of the parameters are adjusted, and each prediction result is finally added to obtain the final prediction result.

Fig. 1. The gradient boosting flow chart of GBDT

Although GBDT has high accuracy, when selecting split points, it needs to traverse each data to calculate the information gain of all possible split points, so this algorithm is time-consuming when processing massive data. Because its computational complexity is proportional to the amount of data and feature dimension On the basis of GBDT, LightGBM adds two new algorithms to solve this problem, namely Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). This not only ensures the high accuracy of the algorithm, but also greatly reduces the consumption of time and storage costs.

3.1 Introduction to the Experimental Data Set

The UK Open University Learning Analytics Dataset (OULAD) was released by the OUAnalyse research group led by Professor ZdenekZdrahal at the Open Data 7 Quintessence Challenge Day on November 17, 2015. The uniqueness of the dataset lies in the richness of the data, which describes student behavior in a virtual learning system. It contains over millions of records generated from 22 courses and more than 32,000 students. The construction of the performance prediction model in this study is centered on the "StudentInfo" worksheet, which is a student information table in the OULAD dataset. This worksheet covers data such as student demographic information and past learning experiences. The features in the dataset provide a more complete description of the student's basic condition and provide a solid foundation for the accuracy of grade prediction. After removing the information that has less relevance to students' grades, such as academic number, we finally get 10 features including "code_module", "gender", "region", "highest_education", "imd_band", "age_band", "num_of_prev_attempts", "studied_credits", "disability", "final_result" are involved in the construction of the MOOC performance prediction model, where final_result is the target variable, and the prediction results will be compared with it to find the error between the prediction results and the true value.

3.2 Preprocessing of the Dataset

In order to make the data meet the usage conditions of the algorithm LightGBM, it is necessary to encode the required features in the dataset. The specific encoding method is shown in Table [1](#page-5-0) below.

Features	Range of values	Code
code_module	[AAA, BBB, CCC, DDD, EEE, FFF, GGG]	$[1-7]$
code presentation	[2013J, 2013B, 2014J, 2014B]	$[1-4]$
Gender	[FA]	[1, 2]
Region	[East Anglian Region, Scotland, North Western Region, [1, 2, 3, 4,, 13] South East Region	
highest education	[HE Qualification, A Level or Equivalent, Lower Than A Level	$[1, 2, 3, \ldots, 5]$
imd band	$[90\% - 100\% , 20\% - 30\% , 30\% - 40\% \ldots]$	$[1, 2, 3, \ldots, 8]$
age band	$[0-35,35-55, > = 55]$	$[1-3]$
Disability	[Yes, No]	[1,0]

Table 1. Feature code table

3.3 Construction of MOOC Performance Prediction Model Based on LightGBM

LightGBM has a faster training speed, lower memory consumption and better accuracy because the algorithm improves the gradient boosting decision tree algorithm in two steps. The former one is the GOSS, and the latter is the EFB. GOSS and EFB can optimize the data processing steps, and then help to reduce the amount of data and the related features. The construction of the LightGBM-based MOOC performance prediction model consists of four core steps, and the details are listed as follow.

Converting Features in a Dataset into a Histogram

Take the feature "imd band" as an example, this feature represents the economic status of the learner. If the previous integrated learning algorithm is used, the economic status of each learner needs to be traversed, and the time overhead required is O (data*feature). Such a time overhead is undoubtedly too high for a data set with a large amount of data. And using the Histogram-based decision tree algorithm can cut the data in a feature into several different bins, and then put the data into the corresponding bins. Therefore, the economic status of the students will be divided into 3 different bins: "1–3", "4–6", and "7–8", and then put the corresponding economic status of each student in the data set into the corresponding bin middle. Then the time over-head will be reduced from the original O(data*feature) to O(feature*bins).

Leaf-wise Algorithm with Depth Limit

LightGBM's leaf growth strategy is a vertical growth strategy, This vertical leaf growth

strategy is called leaf-wise. Leaf-wise will find the leaf with the largest splitting gain from all the current leaves each time, then split, and so on [\[13\]](#page-12-6). For example, after splitting the dataset with the feature "gender $=F$ ", there are two leaves with the features "disability = YES" and "imd_band = 90% -100%", then the algorithm will choose the one with the greater splitting gain for the next split, and the other one will not be split. Therefore, compared with the previous leaf-wise strategy, leaf-wise can reduce more errors and get better accuracy with the same number of splits. However, leaf-wise also has the disadvantage that using this leaf-growing strategy may grow deeper decision trees and produce overfitting. To avoid this problem, the LightGBM algorithm adds a maximum depth limit to the leaf-wise to ensure high efficiency and prevent overfitting at the same time.

Use GOSS to Reduce the Number of Samples

GOSS is a balanced algorithm to reduce the amount of data and ensure the accuracy by excluding most of the samples with small gradients and using only the remaining samples to calculate the information gain [\[13\]](#page-12-6). Taking the feature "imd_band" as an ex-ample, the GOSS algorithm has the following steps.

Sort all samples of the feature "imd_band" to be split according to the gradient size, define two constants a and b. And a represents the value ratio of samples with large gradients, b represents the value ratio of samples with small gradients;

Select a*data_num("imd_band") samples with large gradients, and then select b*data_num("imd_band") samples with small gradients from the remaining samples;

In order not to change the distribution of the data, it is necessary to weight the samples with small gradients and multiply by a constant $\frac{(1-a)}{b}$ to enlarge the weights of the samples with small gradients;

Finally this $(a + b) \times 100\%$ data is used to calculate the information gain.

Use EFB Algorithms for Feature Reduction

High-dimensional data is often sparse, and this sparsity inspires us to design a lossless method to reduce the dimensionality of features. In the LightGBM algorithm, this method of reducing the feature dimension is the mutually exclusive feature bundle EFB [\[13\]](#page-12-6). EFB mainly includes the following steps: ranking the features according to the number of non-zero values;

Calculate the conflict ratio between different features;

Iterate through each feature and try to merge features to minimize the conflict ratio.

The mutually exclusive feature bundle can reduce the time complexity of the histogram from O(#data*feature) to O(#data*bundle), and since the number of bundles is much smaller than the number of features, we are able to greatly accelerate the training process of GDBT without losing accuracy.

4 Experimental Results

4.1 Experimental Setup

Evaluation Metrics

The student performance prediction problem in this experiment is a regression problem,

so the root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) were chosen to evaluate the LightGBM-based MOOC performance prediction results. These three indicators reflect the distance between the predicted value and the actual value, so the closer the values are, the smaller the gap between the predicted value and the groundtruth. And their expressions are shown in Eq. (1) , Eq. (2) , and Eq. (3) , respectively. Where, *n* denotes the number of students, *ypre* denotes the predicted grade of the ith student, and *y* denotes the true grade of the ith student.

$$
\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{pre} - y)^2}
$$
 (1)

$$
\frac{1}{n}\sum_{i=1}^{n}\left(y - y_{pre}\right) \tag{2}
$$

$$
\frac{1}{n} \sum_{i=1}^{n} |(y - y_{p}re)|
$$
\n(3)

Parameter Setting

The parameters setting of the LightGBM are listed in Table [2,](#page-7-3) and it took several iterations of the data using the GridSearchCV function to arrive at this optimal set of parameters. Through practical experiments, it was found that the best results were achieved by using the parameters after tuning the parameters for model construction.

Baseline Methods

The LightGBM algorithm can also work quickly in the face of a huge amount of data, and the prediction accuracy is high. In order to verify the LightGBM-based MOOC performance prediction results, we compared it with the other four machine learning algorithms which were commonly used in the other field, namely Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Adaboost (Adaptive Boosting).

Parameter name	Parameter value	Parameter name	Parameter value
Estimator	gbdt	objective	Regression
learning rate	0.1	n estimators	188
max_depth	4	num leaves	5
max bin	35	min data in leaf	
bagging fraction	0.8	bagging_freq	40
feature fraction	0.6	lambda 11	0.3
lambda 12	0.1	param_grid	params_test5
CV	5	n jobs	-1

Table 2. Parameter setting table

4.2 Experimental Results

Experimental Results

MOOC performance prediction models were constructed based on each of the five algorithms on the jupyter notebook in the Python platform. In terms of the experimental process, the LightGBM algorithm takes the shortest time and occupies the least storage space. In terms of the experimental metric results, the LightGBM algorithm predicts the least error between student grades and final student grades, and the specific performance of each algorithm is shown in Table [3](#page-8-0) below.

	Light GBM	DТ	RF	SVM	Ada
RMSE	0.97	1.34	1.07	0.99	0.98
MSE	0.95	1.81	1.15	0.98	0.96
MAE	0.79	1.02	0.88	0.80	0.81

Table 3. Parameter setting table

This experiment is enough to prove that the LightGBM algorithm has superior performance in the prediction work with large amount of data. Because the three indicators of LightGBM are the lowest, that is to say, the value error predicted by LightGBM is the smallest (see Fig. [2\)](#page-9-0). It shows that GOSS and EFB in the LightGBM algorithm can effectively help MOOC performance prediction to solve the problem of large amount of data and many feature dimensions.

Ranking of Feature Importance

Taking advantage of the interpretability of the LightGBM algorithm, a ranking table of the importance of features affecting the prediction results can be derived during its prediction work (see Fig. [3\)](#page-9-1).

According to this graph, the top 5 factors affecting predicted performance are: "region", "code_module", "studied_credits", "highest_education", and "imd_band". Analyzing the first five factors, "studied_credits" and "highest_education" can be attributed to the learners' existing cognitive structure level, while "region" and "imd_band" can be categorized as the influence of the external environment. Since "code_module" is a factor that cannot be changed objectively, so the next step of this study is to suggest the instructional design of online teaching from two aspects: learners' existing cognitive structure level and external environmental factors.

Analysis of the Advantages of Using Lightgbm for Prediction

According to theoretical analysis and experimental results, the advantages of using Light-GBM for prediction mainly include the following two aspects: First, LightGBM greatly reduces the computational complexity of prediction. This benefits from the GOSS algorithm and the EFB algorithm. They reduce the amount of data and feature dimension. Second, the LightGBM algorithm is interpretable, and this interpretability can help us

Fig. 2. Performance graph of 3 metrics for 5 algorithms

Fig. 3. Map of important features affecting prediction results

find regularities in the data, and these regularities can usually reflect the problems existing in online teaching. Help teachers and students solve problems as soon as possible to improve learning effect.

5 Suggestions for Online Instructional Design Based on Experimental Results

5.1 Different Instructional Designs for Learners in Different External Environments

The learning process of students is a dynamic process, which is affected and restricted by many factors. The external environment is one of the factors affecting students' learning. For example, in the MOOC performance prediction work conducted in this study, "region" and "imd_band" represent the learner's region and the learner's economic status, respectively. These two characteristics, as external environment characteristics, have a great influence on the prediction results. And the mutual information experiment of these two features found a strong correlation between the two features. That is to say, in different regions with different economic conditions, the effectiveness of learners receiving online education is also different. Analyzing the reasons, one cannot be ignored is that the network environment in underdeveloped areas is obviously weaker than that in economically developed areas. The network environment is extremely important for online education, for example, during an epidemic online class, learners with a good network environment can learn without hindrance. However, students in many areas have poor network environments, and they have to spend a lot of time and effort in finding a network environment that can meet the conditions of online classes. Therefore, the delay and lag of the network makes this group of students miss a lot of learning content.

Therefore, according to the different external environments of different learners, the adjusted instructional design can help learners to solve some of the learning difficulties from the external environment. For example, for learners in less developed areas of the network environment, an inquiry-shaped instructional design can be considered, the structure of which is illustrated in Fig. [4](#page-10-0) below. The inquiry-shaped instructional design has strong self-feedback, and learners can obtain information through manipulating media and conduct discovery learning through a series of learning activities such as observation, hypothesis, experimentation, verification, and adjustment. This requires a strong ability to integrate teaching resources and a strong information literacy. In such online teaching, the real-time requirements for the network environment can be

Fig. 4. Illustration of inquiry-shaped teaching structure

appropriately reduced, and the instructor mainly plays the role of guiding and answering questions in the learning process of the learners. Moreover, this way of online teaching allows learners to learn as many times as they want, because the teaching resources developed by the instructor are always available.

5.2 Different Online Instructional Designs for Learners with Different Pre-existing Cognitive Structures

Pre-existing cognitive structure is a key factor in the occurrence and maintenance of learning and is one of the main factors governing the learning process. Important features that affect the prediction results in this study: "studied_credits" and "highest_education", which represent learners' credits and their highest educational level, respectively. Also illustrate the important influence of pre-existing cognitive structures on learners. Before designing an online course, the organizer should understand the learners' existing cognitive structure, which will help to set the starting point of the online course, which is an important prerequisite for tailoring the teaching to the students' needs.

Students are the object of online teaching, and without this element, teaching loses its meaning. In the teaching process, individual differences in student learning are related to the individual student's original knowledge structure, intelligence level, and also to the student's subjective effort. For online teaching organizers, students' intelligence level and effort cannot be easily changed, but they can design different instruction for the same knowledge point according to learners' original cognitive structure. Therefore, it is suggested that online teaching organizers can design different teaching methods and assign different tasks based on learners' education, age, and existing knowledge reserves, so that different learners can reasonably choose the online learning courses according to their actual situations.

Acknowledgements. This work is supported by Yunnan Normal University Graduate Research Innovation Fund (Grant No. YJSJJ22-B88), Yunnan Innovation Team of Education Informatization for Nationalities, and Scientific Technology Innovation Team of Educational Big Data Application Technology in University of Yunnan Province.

References

- 1. Luo, Y., Xibin, F., Han, S.: Exploring the interpretability of a student grade prediction model in blended courses. Distance Educ. China, 46–55 (2022)
- 2. Xian, Wei, F.: The evaluation and prediction of academic performance based on artificial intelligence and LSTM. Chinese J. ICT Educ. 123–128 (2022)
- 3. Hao, J.F., Gan, J.H.: MOOC performance prediction and personal performance improvement via Bayesian network. Educ. Inf. Technol. 1–24 (2022)
- 4. Hasan, F., Palaniappan, S., Raziffar, T.: Student academic performance prediction by using decision tree algorithm. In: IEEE 2018 4th International Conference on Computer and Information Sciences (ICCOINS), pp. 1–5 (2018)
- 5. Zhao, H.Q., Jiang, Q., Zhao, W., Li, Y., Zhao, Y.: Empirical research of predictive factors and intervention countermeasures of online learning performance on big data-based learning analytics. e-Educ. Res. 62–69 (2017)
- 6. Li, S., Li, R., Yu, C.: Evaluation model on distance student engagement: based on LMS data. Open Educ. Res. **24**(01), 91–102 (2018)
- 7. Wei, S.F.: An analysis of online learning behaviors and its influencing factors:a case study of students' learning process in online course open education learning guide in the open university of China. Open Educ. Res. 81–90+17 (2012)
- 8. Qing W, Ru-guo L.: Predicting the students' performances and reflecting the teaching strategies based on the e-learning behaviors. Modern Educ. Technol. 6, 18–24 (2017)
- 9. Xu Xiaoyu, F.: Research on the prediction and early warning model of student achievement based on heterogeneous information network. Inf. Technol. Netw. Secur. 84–89 (2022)
- 10. Márquez, C., Vera, F.: Early dropout prediction using data mining: a case study with high school students. Expert Syst. **31**(1), 107–124 (2016)
- 11. Lykourentzou, F., Giannoukos, S., Nikolopoulos, T.: Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput. Educ. **53**(3), 950–965 (2009)
- 12. Ke, G., Qi, F., Thomas, M.S., Finley, T.: LightGBM: A Highly Effificient Gradient Boosting Decision Tree. Curran Associates Inc, Neural Information Processing Systems (2017)
- 13. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. **30** (2017)