Student Achievement Analysis and Prediction Based on the Whole Learning Process

Meixue Wu, Hong Zhao*, Xiaoyu Yan, Yun Guo, Kai Wang
College of Computer Science
Nankai University
Tianjin, China
zhaoh@nankai.edu.cn, meixuew@mail.nankai.edu.cn

Abstract—**Blended learning is increasingly used in college teaching, and formative evaluation has become the main method for assessing student performance. Based on the formative evaluation data of an existing course, how to model, analyze and predict the possible problems of students in the future learning process and give recommendations on learning strategy are problems worthy of in-depth study. In this paper, Apriori algorithm was used to perform association analysis on the formative evaluation data of the Fundamentals of Programming course in Nankai University, the results indicate that there are strong association rules between SPOC video scores, case study assignments scores, etc. K-Means algorithm was used to perform cluster analysis on SPOC platform scores, offline course scores and final exam scores, the results indicate that the advantages and disadvantages of students of different categories are consistent in two semesters. Finally, the clustering results of the first semester were added to the data set, Random Forest was used for feature selection, and four ensemble learning models were trained respectively to predict final exam grades. The results show that the XGBoost model works best, the accuracy of predicting the final exam grades of two semesters is 77.02% and 80.10%, respectively.**

Keywords—blended learning, association rules, cluster analysis, ensemble learning

I. INTRODUCTION

In recent years, the emergence of blended learning pedagogy is rapidly changing the traditional mode of teaching and learning. Some studies have shown that blended learning has better learning effect than traditional teaching approach [1], [2]. There have been many studies and cases that used blended learning data to mine and analyze student performance. Lu et al. [3] made early predictions of students' final academic performance in a blended calculus course. This study collected real data from 21 variables in the blended course, and applied principal component regression to predict students' final academic performance. Zacharis et al. [4] used CART decision tree to classify the students and predict who is at risk based on the data from four online activities (message exchange, creating wiki content groups, opening course files, and online testing). Shin et al. [5] used clickers to record attendance, quiz performance, and perform daily surveys. Their results show that the blended learning method improves student achievement levels. Sukhbaatar et al. [6] used decision tree to identify dropout prone in the middle of semester. The data included 717 sophomores' online activities in compulsory course. Gabrijela et al. [7] performed association analysis in a blended learning environment to improve the examination process. They've discovered a large number of association rules through Apriori algorithm, and their results enabled teachers to better understand the concept of creating tests and decide how to improve the test design. Bratislav et al. [8] conducted a comparative analysis of predictive classifiers in a blended learning environment. Using different classifiers(Naive Bayes, Hidden Naive Bayes, etc.) to obtain important results in different categories, and then using the majority voting scheme to form an ensemble based on Naive Bayes, Hidden Naive Bayes, J48 decision tree and Random Forest.

The Fundamentals of Programming offered by Nankai University is divided into Fundamentals of Programming (Part 1) and Fundamentals of Programming (Part 2), which are respectively offered in two semesters in one academic year. This course is a public compulsory course, students from 8 majors including mathematics, chemistry, history, etc. This course adopts blended learning mode and the curriculum design could be summarized as following:

- Before class, students watch the teaching videos on SPOC (Small Private Online Course) platform, participate in discussions in the discussion board, etc. This part is mainly based on students' autonomous learning;
- In class, teachers give quizzes to check the effectiveness of students' self-learning, and explain in-depth the key points, difficult points, as well as the error-prone points discovered from the quizzes;
- After class, students finish case study assignments, problem solving project, etc.;
- Final exam accounts for 30% of the total score.

The research data of this paper is from the two-semester course, Fundamentals of Programming. The participating objects, teaching process and evaluation methods of the two teaching cycles are consistent, and formative evaluation data is collected from each stage of the blended course.

For most of the research on blended learning, data used to do data mining was from one course of one semester, or different courses of multiple semesters. It is rare to study the data of one course with two semesters. In two semesters with the same teaching process, whether the students' performance tends to be the same or there are obvious differences is what we are interested in and worth further study. Therefore, this study explored the learning behavior of students by analyzing and comparing all the teaching data of a two-semester course.

II. METHODOLOGY

A. Data Collection

This study was based on data from two semesters of 368 first-year students who took the Fundamentals of Programming course at Nankai University. A student's total

978-1-7281-7267-5/20/$31.00 ©2020 IEEE

score of each semester consists of scores on SPOC platform, scores in offline course and final exam score. Scores on SPOC platform includes SPOC video score, SPOC quiz score, SPOC homework score and SPOC exam score; scores in offline course include case study assignments score, problem solving project score and attendance score. The data participating in this study is shown in Table I. In addition to the course scores, they also include students' gender, major and province. Student performance in the college entrance examination was also considered as a research feature, since to some extent, it reflects the ability of students at the time of enrollment.

TABLE I. EXTRACTED FEATURES

Category	Feature
Identity	Gender, Major, Province
Score	*SPOC platform*: SPOC video score, SPOC quiz, SPOC homework score, SPOC exam score
	Offline course: case study assignments score, problem solving project score, attendance score
	Final exam score
	College Entrance Examination Score

B. Data Processing

1) College entrance examination score

Since students came from different provinces, the full mark of the college entrance examination varies from province to province, the college entrance examination scores need to be processed. Suppose that a student's entrance examination score is qS, and the full mark in the student's province is TQS, then this student's processed college entrance examination score (qs) is:

$$qs= (qS/TQS)*100 \qquad (1)$$

2) Mapping course scores to grades

All the scores involved in this paper were mapped to 5 levels, taking the full mark of 100 as an example, 60 is the passing mark, below 60 is unqualified. Above 60, every 10 points belongs to a grade, which are altogether 5 grades, A, B, C, D and E. Since the original full mark of each score is different, it is divided by percentage. The mapping relationship between score and grade is shown in Table II. Where S is the original score and TS is the full mark.

TABLE II. THE MAPPING RELATIONSHIP BETWEEN SCORE AND GRADE

Grade	Score
A	$TS*0.9 \le S \le TS$
B	$TS*0.8 \le S < TS*0.9$
C	$TS*0.7 \le S < TS*0.8$
D	$TS*0.6 \le S < TS*0.7$
E	$0 \le S < TS*0.6$

3) Dividing students' scores

Scores on the SPOC platform reflect students' learning achievements on the SPOC platform, scores in offline course reflect students' participation in offline course and their completion of daily homework, what's more, the final exam examines the students' mastery of knowledge at the end of semester. Therefore, students' scores in this study were calculated into the following three scores:

- SPOC platform score, which equals the sum of the four original scores of SPOC video, SPOC quiz, SPOC homework and SPOC exam;
- Offline course score, which equals the sum of the three original scores of case study assignments, problem solving project and attendance;
- Final exam score, which is the original score of the final exam.
- Finally, SPOC platform scores, offline course scores and final exam scores are normalized.

C. Association Rules

Association rules mining is to find possible associations or connections between things from the data [9]. Apriori algorithm is one of the most influential classical algorithms for mining frequent itemsets [10], [11], K-1 order itemsets are used to search K order itemsets. It consists of two subtasks: frequent itemsets mining and strong association rule determination. The Apriori algorithm is used to build frequent itemsets. The relevant definitions of association rules are as follows:

- Association rule: When both X and Y are itemsets, itemset X and itemset Y have the following relationship for transaction set S: both X and Y are subsets of S, and the intersection of X and Y is empty. An association rule looks like X→Y, where X and Y called antecedent and subsequent of association rule respectively.
- Confidence: Then the confidence of X→Y is the probability of Y occurring simultaneously in the event that X occurs.
- Support: The support of X→Y is the probability that X and Y appear at the same time.
- Strong association rule: Set the minimum confidence and minimum support of association rules. If the confidence and support of rules are not less than the minimum confidence and minimum support, they are called strong association rules.

In this study, Apriori algorithm was used to mine the association rules among course grades, in order to find the potential correlation between grades in the two semesters.

D. Clustering

Cluster analysis is an exploratory and unsupervised data analysis method. Cluster analysis can classify data, the data with high similarity can be classified into the same category, while the data with low similarity can be classified into different categories [12]. The measurement of the degree of similarity is the distance between two data elements. K-Means is an iterative cluster analysis algorithm. It's steps are as follows:

Step 1: Select an initial cluster center for each cluster;

Step 2: Allocate the sample set to the nearest cluster according to the minimum distance principle;

Step 3: Update the cluster center using the sample mean of each cluster;

Step 4: Repeat step 2 and step 3 until the cluster center no longer changes;

Step 5: Output the final cluster center and k cluster divisions.

In order to analyze students' performance in online course, offline course and final exam in two semesters, K-Means was used to cluster the SPOC platform scores, offline course scores and final exam scores. Finally, the clustering results were presented in a chart to visualize the performance characteristics of students in different categories.

E. Prediction

1) Feature engineering

The purpose of feature selection is to analyze the effectiveness of each feature from a theoretical perspective and select the most representative feature subset with the best classification performance to effectively describe the input data, it is an important task before building prediction model. In this paper, Random Forest was used to sort the feature importance.

2) Prediction model

The maximum number of features in this paper is 20. For data with lower feature dimensions, feature engineering plus traditional machine learning methods will have stronger generalization ability, and the results are also more interpretable and adjustable. Ensemble learning in machine learning is to combine several weak supervised models in order to get a better and more comprehensive strong supervised model. The underlying idea is that even if one weak classifier gets a wrong prediction, other weak classifiers can correct the error back. Therefore, four kinds of popular ensemble learning algorithms are selected in this study to establish the final exam grade prediction model.

- RF(Random Forest). An algorithm that uses multiple decision trees to train and predict samples.
- AdaBoost. With the idea of iteration, only one weak classifier is trained in each iteration, and the trained weak classifier will participate in the next iteration.
- GB(Gradient Boosting). The goal of each weak classifier is to fit the negative gradient of the loss function of the previous accumulative model, so that the cumulative model loss after adding the weak learner can be reduced in the direction of the negative gradient.
- XGBoost. It is based on CART tree, with negative gradient as the learning strategy.

The final exam is the last part of the Fundamentals of Programming course, and the review time will be reserved for the students before the exam. In order to give early warning to students and let them adjust the review pace more reasonably, in this paper, students' identity characteristics and grades obtained in the course were used to predict the final exam results of each semester. The clustering results of the first semester were added into the prediction of the second semester as an important feature.

3) Evaluation indicators

For the classification task, accuracy, precision, recall and F1-score were used as evaluation indexes. In this paper, the final classification label is A, B, C, D and E, which are multi-classification task. The calculation formulas of these four evaluation indexes are as follows:

$$\text{Accuracy}=\sum_{i=1}^{N}\frac{TP_i}{TP_i+TN_i+FP_i+FN_i} \tag{2}$$

$$\text{Precision}(P)=\frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FP_i} \tag{3}$$

$$\text{Recall}(R)=\frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FN_i} \tag{4}$$

$$\text{F1}-\text{score}=\frac{2*P*R}{P+R} \tag{5}$$

Where N represents the number of categories, TP_i, TN_i, FP_i and FN_i represent true positive, true negative, false positive and false negative for the *i*th category, respectively.

The overall framework of this study is shown in Fig. 1. Firstly, data was collected and preprocessed. Secondly, association analysis and clustering analysis were carried out. Finally, prediction models to predict final exam grades were established. In the project of establishing prediction model, feature selection was carried out first. Since the clustering results of students in the first semester had become known terms when the prediction model of the second semester was established, the clustering analysis results of the first semester were also taken as a feature to be selected together with the original data set.

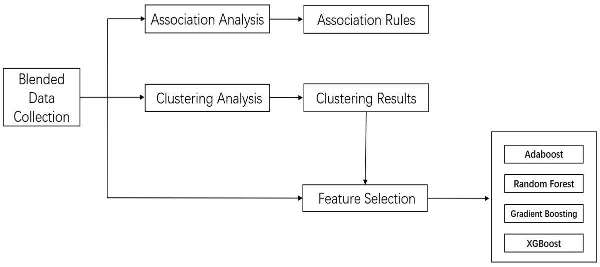

Fig. 1. Overview of the work.

III. RESULT

All experiments in this paper were written in Windows 10 environment using Python 3.6.

A. Association Rules Mining

The Apriori algorithm was used to mine the association rules of grades in two semesters to explore the potential relationship between grades. Set the minimum confidence to 0.75 and minimum support to 0.65. Table III shows the mapping relationship between the full name and abbreviation of grades involved in the association rules. The association rules are shown in Table IV. In Table IV, the symbolic meaning of association rules is: F(First) represents the First semester, S(Second) represents the Second semester, and A is a grade. If an association rule is X-Y, then confidence 1 is the confidence of X→Y and confidence 2 is the confidence of Y→X.

TABLE III. THE MAPPING RELATIONSHIP BETWEEN FULL NAME AND ABBREVIATION

Full Name	Abbreviation
SPOC video	video
SPOC exam	exam
case study assignments	case
attendance	attend
problem solving project	prosolve

TABLE IV. THE RESULTS OF ASSOCIATION ANALYSIS

	Association rules	Confidence 1	Confidence 2	Support
1	caseFA-videoFA	95.9%	80.1%	75.7%
2	examFA-videoFA	90.8%	79.3%	74.9%
3	attendFA-videoFA	91.4%	87.3%	72.6%
4	attendFA-caseFA	80.4%	76.4%	70.3%
5	caseSA-videoSA	89.1%	96.3%	74.7%
6	attendSA-videoSA	88.9%	86.4%	76.0%
7	attendSA-caseSA	80.2%	85.7%	71.5%
8	videoFA-videoSA	89.0%	95.7%	74.2%
9	attendFA-attendSA	89.8%	87.8%	76.8%
10	prosolveFB-prosolveSB	82.7%	75.6%	65.3%

Taking the first rule as an example, caseFA represents the case study assignments grade of A in the first semester, videoFA represents the SPOC video grade of A in the first semester. Confidence 1 of 95.9% means that among the students who got A in the first semester SPOC exam, 95.9% of them got A in the first semester SPOC video too, and Confidence 2 of 80.1% means that 80.1% of the students who got A in the first semester SPOC video also got A in the first semester case study assignments. Support of 75.7% means that 75.7% of the students got A both in case study assignments and SPOC video in the first semester. The first three association rules are all about the first semester, from which we can see that students who got A in SPOC video, highly likely got A in SPOC exam, case study assignments and attendance. The fourth association rule shows that in the first semester students' who got A in attendance had a 80.4% chance to got A in case study assignments. Association rules 5 to 7 are all about the second semester and the same with the first semester. Association rules 8 and 9 are about SPOC video grades and attendance grades, these two respectively reflect students' participation in online and offline classrooms. In combination with the 3, 6, 8, 9 association rules, it can be concluded that students who actively participated in the SPOC platform were most likely to actively participated in offline course. Students who actively participated in the course in the first semester highly likely had the same positivity in the second semester. The last association rule shows that students who got B in problem solving project in the first semester highly likely got B in the second semester too.

B. Cluster Analysis

The first semester's SPOC platform scores, offline course scores and final exam scores were clustered. Take K = 3, the clustering results are shown in Fig. 2. F represents the first semester.

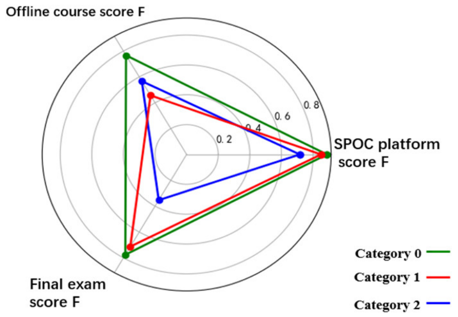

Fig. 2. Clustering results of the first semester.

Category 0: Students of this category were excellent in all three aspects. They actively participated in the learning on SPOC platform, earnestly completed the homework assigned by teachers, and achieved excellent results in the final exam. Such students account for 43% of the total number.

Category 1: Students of this category had excellent scores on SPOC platform and final exam, but offline course scores were low. Such students actively studied through the SPOC platform and attached importance to the final exam, but they performed poorly in offline course. Such students account for 35% of the total number.

Category 2: Students of this category had relatively low scores, especially the final exam scores. Such students account for 22% of the total number.

The second semester's SPOC platform scores, offline course scores and final exam scores were clustered. The clustering results are shown in Fig. 3. S represents the second semester.

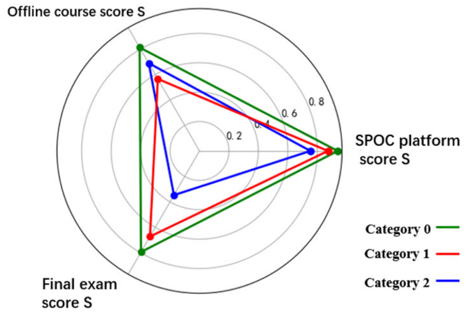

Fig. 3. Clustering results of the second semester.

The clustering results of the second semester are basically similar to those of the first semester, and the proportion of the three categories of students in the total population is 47%, 29% and 24%, respectively.

In order to analyze the characteristics of students of different categories in two semesters, scores in two semesters were clustered. Take k=4, the clustering results are shown in Fig. 4.

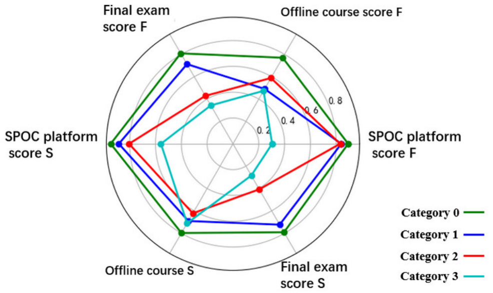

Fig. 4. Clustering results of two semesters.

Category 0: Students of this category performed excellent and balanced in all aspects, with no obvious weaknesses. Such students account for 35% of the total number.

Category 1: Students of this category scored lower in two semesters of offline course, and performed excellent in other items. Such students account for 32% of the total number.

Category 2: Students of this category scored lower in two semesters' final exam, and performed moderate in other items. Such students account for 25% of the total number.

Category 3: Students of this category had poor performance in all aspects, especially in the final exam. Moreover, they scored lower on SPOC platform compared with those of other three categories. Such students account for 8% of the total number.

It can be found that the clustering results in two semesters are basically the same, and when the data of the two semesters were clustered together, the advantages and disadvantages of each category in the two semesters are basically the same. In addition, most students achieved excellent results on the SPOC platform, while the final exam was challenging for many students.

C. *The Prediction Model of Final Exam*

1) *Final exam grade prediction of the first semester*

Firstly, a total of 11 features in the first semester performed feature selection using Random Forest. Features include grades of the first semester, students' identity information and college entrance examination score(qs). The results of feature importance ranking are shown in Fig. 5. F represents the first semester.

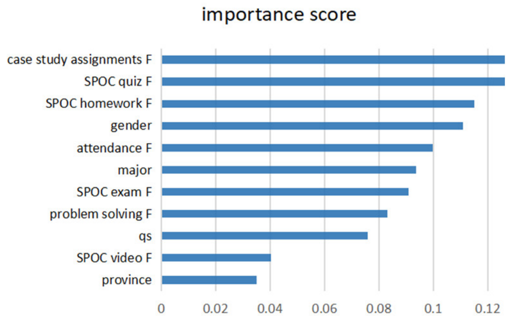

Fig. 5. The ranking of feature importance in the first semester.

According to the ranking of feature importance, discarding the last two features, province and SPOC video grades, the best results of final exam grade prediction were obtained by using the remaining 9 features. The algorithms used were Adaboost, Random Forest, Gradient Boost and XGBoost. The model called functions in the scikit-learn (sklearn) module, and the data set was divided into a 70% train set and a 30% test set. The iteration times of the model were 100, 50, 200 and 200 respectively. The prediction results are shown in Table V.

TABLE V. THE FIRST SEMESTER PREDICTION RESULTS

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Adaboost	66.8	66.9	65.9	66.4
RF	73.1	73.8	69.6	71.6
GB	76.1	76.3	74.6	75.4
XGBoost	77.6	77.5	76.3	76.9

2) *Final exam grade prediction of the second semester*

The K-Means clustering results of the first semester were known data for the second semester, so added the results to the data set, then Random Forest was used to rank the feature importance degree of 20 features. Except 11 features in the first semester, it also includes clustering results of the first semester(category F) and grades in the second semester. The results of feature importance ranking are shown in Fig. 6. S represents the second semester.

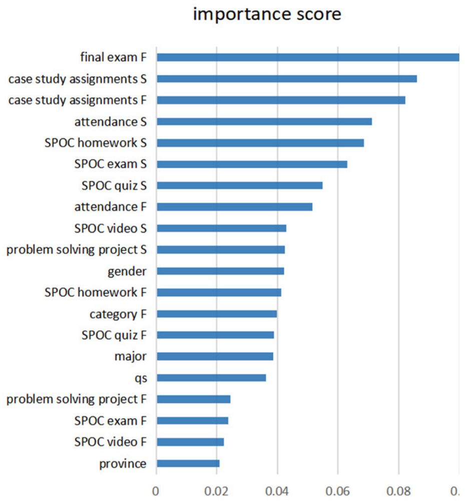

Fig. 6. The ranking of feature importance in the second semester.

By feature selection results, discarding problem solving project grades(F), SPOC exam grades(F), SPOC video grades(F) and province, using the rest of the 16 features to establish the second semester's final exam grades prediction model got the best effect. Model parameters and data set partitioning same with the first semester. The prediction results are shown in Table VI.

TABLE VI. THE SECOND SEMESTER PREDICTION RESULTS

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Adaboost	71.6	74.0	66.4	70.0
RF	74.4	75.9	74.4	75.1
GB	77.5	78.8	78.1	78.4
XGBoost	80.1	78.5	79.6	79.0

The best performing algorithm is XGBoost, with the highest prediction accuracy of 77.6% and 80.1% for the two semesters respectively, which are good results for multi-classification tasks and can provide helpful warning and guidance for learners and teachers.

IV. Conclusion And Discussion

This study explored the characteristics of student achievement in a two-semester blended learning course. Based on the formative evaluation data of the course, Apriori algorithm was used to mine association rules between various grades in two semesters. K-Means algorithm was used to classify students and discuss the performance characteristics of different categories. Finally, machine learning algorithms were used to establish final exam grade prediction models.

The results of association analysis show that students who actively participate in online and offline class in the first semester will also keep an active attitude of participation in the second semester. The results of clustering show that for each category of students, their relative advantages and disadvantages among the SPOC platform scores, offline course scores and final exam scores remain the consistent in two semesters. Overall, most students' SPOC platform performance was outstanding, only 8% of students got bad SPOC platform scores, but for some students the final exam was challenging. The possible reason is that the teaching video on SPOC platform could be watched repeatedly, the quiz was relatively simple, but the final exam was a closed-book exam, which measured students' memory and mastery of knowledge. Through the above association analysis and clustering analysis of the Fundamentals of Programming course, it can be found that under the blended learning mode, most students will show similar learning effects in two semesters for a course with a one-year learning cycle.

Finally, by combining the characteristics of students' identity, course grades and clustering results, models that can effectively predict the final exam grades were established by using machine learning. In the feature selection of the first semester, the SPOC video was less important, which may be because the video can be played repeatedly, even if you do not watch it carefully, there is still a play record. In the feature selection of the second semester, the final exam of the first semester was the most important feature. It indicates that the final exam scores of the two semesters have a certain degree of correlation. And the problem solving project, SPOC video and SPOC exam in the first semester ranked low in the second semester, which may be because these three characteristics ranked low in the first semester, so they had less impact on the final exam grades of the second semester. It can be seen from the model training that the model accuracy in the second semester (80.1%) is significantly higher than that in the first semester (77.6%). It is mainly because the modeling of the second semester used two semesters' features, the final exam grades of the first semester ranked first in importance when predicting final exam results of the second semester, and the clustering results of the first semester had an important influence on the prediction of the second semester.

In the future, it is necessary to establish and improve the prediction and feedback mechanism of the results under the blended learning mode, analyze the results of students after the end of the first semester, provide feedback on the learning effect for teachers and students, and remind students to adjust their learning strategies as soon as possible. Secondly, it's worth consideration to further strengthen the data collection function of SPOC platform, such as the click amount recording of each teaching video, which is conducive to more detailed teaching data analysis in the future.

Acknowledgment

The authors gratefully acknowledge the research funding from the following programs:

- The Key Cultivation Project of Education Achievement Award of Tianjin (PYGJ-018)
- Teaching reform of Nankai University(NKJG2020150)

References

[1] A. Padilla-Meléndez, A. R. Del Aguila-Obra, and A. Garrido-Moreno,"Perceived playfulness, gender differences and technology acceptance model in a blended learning scenario," Comput. Educ., vol. 63, pp. 306–317, Apr. 2013.

[2] C. J. Asarta and J. R. Schmidt, "Access patterns of online materials in a blended course," Decision Sci. J. Innov. Educ., vol. 11, no. 1, pp. 107–123, 2013.

[3] Owen H. T. Lu, Anna Y. Q. Huang, Jeff C.H. Huang, Albert J. Q. Lin, Hiroaki Ogata and Stephen J. H. Yang. "Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning." Journal of Educational Technology & Society, vol. 21, no. 2, pp. 220–232, 2018.

[4] Nick Z. Zacharis, "Classification and Regression Trees (CART) for Predictive Modeling in Blended Learning," International Journal of Intelligent Systems and Applications(IJISA), vol. 10, no. 3, 2018.

[5] Youhyun Shin, Junghyuk Park, Sang-goo Lee. "Improving the integrated experience of in-class activities and fine-grained data collection for analysis in a blended learning class". Interactive Learning Environments, vol. 26, pp. 1-16, 2018.

[6] O. Sukhbaatar, K. Ogata and T. Usagawa, "Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), pp. 2205-2208, 2018.

[7] Gabrijela Dimić,Bratislav Predić,Dejan Rančić,Vera Petrović,Nemanja Maček,Petar Spalević, "Association analysis of moodle e ‐ tests in blended learning educational environment," Computer Applications in Engineering Education, vol. 26, pp. 417-430 2018.

[8] Bratislav Predić,Gabrijela Dimić,Dejan Rančić,Perica Štrbac,Nemanja Maček,Petar Spalević, "Improving final grade prediction accuracy in blended learning environment using voting ensembles," Computer Applications in Engineering Education, vol.26, 2018.

[9] CHENG Xue-Qi, JIN Xiao-Long, WANG Yuan-Zhuo, GUO Jia-Feng, ZHANG Tie-Ying and LI Guo-Jie, "Survey on big data system and analytic technology," Journal of Software, vol. 25, pp. 1889-1908, 2014.

[10] AGRAWAL R, SRIKANT R. "Fast algorithm for mining association rules," Processdings of 20th Int. Conf. VeryLarge Data Bases(VLDB).Morgan KaufmanPress, pp. 487-499, 1994.

[11] Xing Changzheng , Anweiguo , Wang Xing. "Improvement of algorithm for mining frequent itemsets in vertical data format," Computer Engineering and Science, vol.39, pp. 1365-1370, 2017.

[12] Yu Qilin. "Optimization of initial clustering center selection for K-means algorithm," Journal of Computer System Applications, vol. 26, pp. 170-174, 2017.