# eduDiag

## Machine learning visual diagnosis based on student behavior and performance prediction

/Tingting Zhang    /Jiansu Pu    /Yuwei Zhang    /Yulu Xia    / Haixing Dai    /Jingwen Zhang    /Hui Shao    / Shaolun Ruan
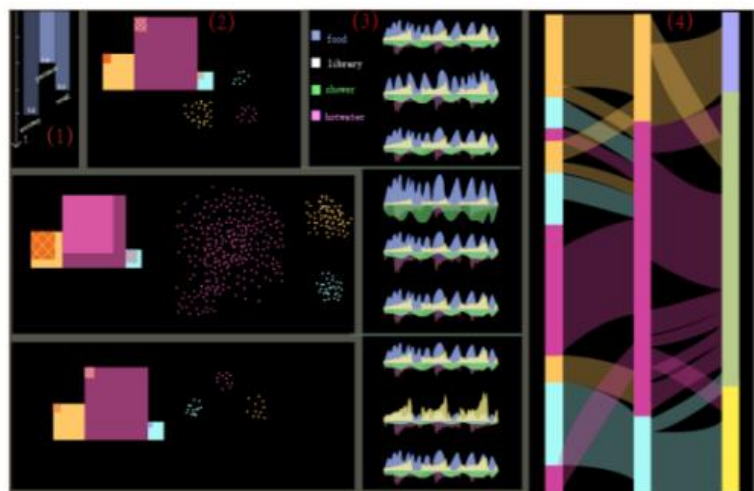


Figure 1: (1) prediction module, (2) prediction classification module, (3) student behavior module and (4) academic performance module

**Intruduction** Interactive model analysis, the process of understanding, diagnosing, and refining a machine learning model with the help of interactive visualization, is very important for users to efficiently solve real-world artificial intelligence and data mining problems [1,3].The measurement of students' school performance plays an important role in the evaluation of educational quality.

## VISUALIZATION  DESIGN  SYSTEM

**IMPLEMENTATION**   The system consists of four modules :(1) prediction results module, (2) prediction classification module, (3) student behavior module and (4) academic performance module. The prediction results module shows the final accuracy, recall and accuracy of the performance prediction with random forest. In the prediction classification module, the collected prediction data are presented and compared with the data of different categories.    The student behavior module shows the behavior patterns of students entering and leaving the library, consuming food in the cafeteria, bathing and fetching water in the dormitory, and the data will be used as the input characteristics of the random forest. The academic performance module is to draw the trend of students' performance predicted to be different from each other .

**CASE  STUDY** In this study, analysis of consumption data in smart card data, library check-in data (including time, location, and encrypted student ids) was associated with academic records extracted from students' final gpa from 2009-2012. 2. The following conclusions: Let's look at module (1) first. The prediction accuracy of the model reached 80%, but the recall and accuracy of the model were not very good, only 60% and 40%.See module (2), the inside of the said three rectangular size according to the classification of grades A, B, C three kinds of the size of the number of students, we can see that most class B students, account data model class imbalance problems, and then looked at the second line, it is by mistake classification for the prediction of class B, A, C in many data have been misclassification in class B, this is because in the learning process of the model, , model A strategy that in order to guarantee the improvement of accuracy, in order to make models don't learn, so I need to solve the problem of unbalanced class. and to the characteristics of different weights, Modules (3) and (4) show the behavior patterns and academic performance trends of students.Thus. In order to improve the accuracy of the model, weight should be added to the behavior characteristics.
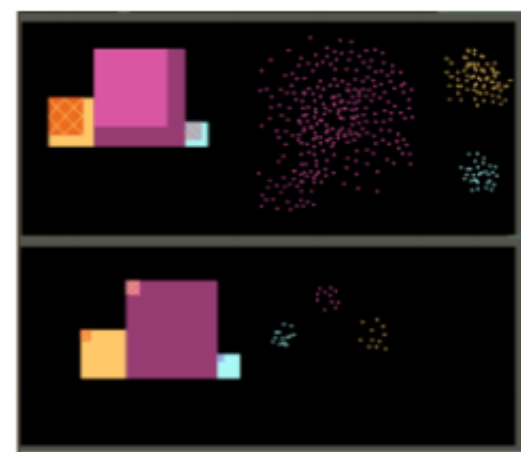


Figure 2 : prediction classification module

**CONCLUSION**    Due to time and energy constraints, there is still much room for improvement in our work.    Through the case study, we found that the algorithm combined with time series can help us predict students' performance. However, the data set is sparse, it is not good for the prediction results of the model.In the future, we plan to design better methods to divide the data, better solve the class imbalance, and strive for more satisfactory model results.