The student grade prediction project aims to predict whether students can get a passing grade (70 points or above) by analyzing students' behavioral data (such as the number of questions asked, the number and accuracy of answers, the number of comments, etc.)

Data description:

Questions\_Authored The number of questions asked by students; Answers\_Submitted The number of questions answered by students; Answers\_Correct The number of correct answers given by students; Comments\_Written The number of student comments; Total\_Character\_Count The total number of characters in student comments; Student\_Avg\_Score The average quality score of the questions asked by students (i.e., judging whether the question raised by the student is good or not); Grade Student grade; Above\_70 According to Grade, if the student's grade is greater than or equal to 70, it is 1, and if it is less than 0.

Use the first 6 variables to model and predict whether the student's grade can be greater than or equal to 70 points. It is hoped that the prediction accuracy of both 0 and 1 categories will reach more than 70%.

This project involves a binary classification task in machine learning. The goal is to predict whether their final grade is qualified through students' activity data (such as questions, answers, comments, etc.).

Models used: svm, random forest, gradient boosting machine (GBM), DNN (Deep Neural Network), KNN (K-Nearest Neighbors), LSTM, MLP (Multi-Layer Perceptron)

All the above models can be used for prediction of classification tasks.

Evaluation indicators: MSE, Precision, recall, specificity, roc AUC score

And output the visualization image of the confusion matrix

Then conduct model comparison analysis.

Finally, it was found that DNN had the best effect, with a prediction accuracy of up to 70%, while the remaining LTSM, MLP, and random forest also had good effects, reaching more than 65%. However, for SVM and GBM, it was only about 60%. And through the Bayesian optimization algorithm, the effects of DNN and LSTM models can be further improved to achieve a prediction effect of 75%.

And through key factor analysis, it was found that the main factors affecting students' grades are ranked in order of relevance as follows: Student\_Avg\_Score, the average quality score of questions asked by students (that is, judging whether the questions asked by students are good or not), Answers\_Correct, the number of correct answers given by students, and Answers\_Submitted, the number of questions answered by students.