Student Grade Prediction Model Based on RFE_RF and Integrated Learning Voting Algorithm

Yajing Niu, Tao Zhou(✉), Zhigang Li, and Haochen Liu

College of Information Science and Technology, Shihezi University, Shihezi 832000, China
{zt_inf,Lzg_inf}@shzu.edu.cn

Abstract. As an important branch of educational big data, grade prediction has become a hot spot for researchers. In order to predict students' grade levels more accurately and have good prediction accuracy at each grade level, the RFE _ RF feature selection method is proposed to reduce the dimension of features. Several machine learning models, such as the decision tree, random forest, logistic regression, and Naive Bayes, are used to construct a weighted voting model based on information entropy to build a student grade prediction model with an accuracy rate of 84.38%. Compared with the performance of other single models, the accuracy, F1-score and recall rate of this model are all good at low, middle, and high-grade levels. The results show that the algorithm can provide a reference for the study of grade-influencing factors and student grade prediction model.

Keywords: Feature selection · information entropy · voting algorithm · student grade prediction

1 Introduction

In recent years, the Ministry of Education, based on the era of informationization 2.0, issued the Notice on Organizing the Declaration of "5G + Smart Education" Application Pilot Project to encourage the integration of education and technology with data as the foundation and to innovate the building of an education platform with scientific decision-making, refined management process, and timely teaching analysis from three levels of data collaboration, data refinement and data openness [1, 2].

In the field of education, the combined application of big data involves many aspects, mainly including two technologies, i.e., educational data mining and learning analysis, both of which provide important support for improving school planning, increasing teaching efficiency, enhancing learning quality, and accelerating interdisciplinary development [3]. These two technologies also have different applications in the field of teaching and learning. For example, Andres Gonzalez-Nucamendi [4] used questionnaires to collect characteristics of engineering students in various dimensions such as enthusiasm for independent learning, so as to identify the relationship between independent and dependent variables and predict students' grades; Ding [5] performed cluster analysis on

© The Author(s) 2023
C. F. Peng et al. (Eds.): EIMT 2023, AHSSEH 8, pp. 1193–1201, 2023.
https://doi.org/10.2991/978-94-6463-192-0_154

students by subdividing students' behaviors into different groups from various aspects based on students' characteristics in the big data of the management system of Xi'an University of Technology; Ahmed A [6] using CONV-LSTM algorithm constructed a student dropout prediction model based on seven behavioral data such as students viewing pages in MOOC to identify students at risk of dropping out of school.

The research on course grade prediction is mainly to select the features with higher influencing factors as input features according to different learning situations, and select appropriate machine learning algorithms to construct the optimal prediction model to predict the student grade prediction model. For example, Hangjie Shen [7] used two algorithms, fuzzy clustering and support vector regression, to predict students' final grades using student grades and parents' education level as characteristics. Jakub Kuzilek [8] identified students who failed the exam in the first academic year with four machine learning models to help teachers identify and intervene with at-risk students.

From the above research, it can be concluded that with different features and algorithms selected in different learning situations, the final prediction results are also different. We propose a random forest algorithm based on feature recursive elimination and an integrated voting algorithm, which not only provides an effective method for teachers to carry out personalized teaching and predict course grades, but also offers reasonable suggestions and opinions for future teaching focus.

2 Data Preprocessing Algorithms

The original dataset usually contains some features that are less relevant to the final grades, and the accuracy of the model will be improved by removing these irrelevant attributes. We use the RFE_RF feature selection method combining recursive feature elimination (RFE) and random forest (RF) to determine the optimal subset by calculating the root mean square error of the model, and add 5-fold cross-validation to the outer layer of the algorithm for better performance evaluation and feature selection basis. The selection process of the feature selection method includes input and output processes. The specific algorithm is as follows:

Input training samples $\mathcal{F}\{\mathcal{X}_k, \mathcal{Y}_k\}$ (k = 1,2,3,…,n), $\mathcal{Y}_k \in$ {low-level interval, mid-level interval, high-level interval}.

Output feature ordering set.

Step 1: Initialize the feature set $\mathcal{F}$ as the original data set and the feature ordering set as empty.

Step 2: Loop the following process until $\mathcal{F}$ is empty.

1) Obtain the training samples of the feature set to be candidates.

2) Generate a decision tree after $\mathcal{F}$ is resampled by bootstrap to build a random forest classification model, get the final classifier results after final voting, and calculate the importance of each input feature to the model output. The importance index is calculated by the following formula (1).

$$P_k = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t} D_{Gkij}}{\sum_{k=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{t} D_{Gkij}} \times 100\% \tag{1}$$

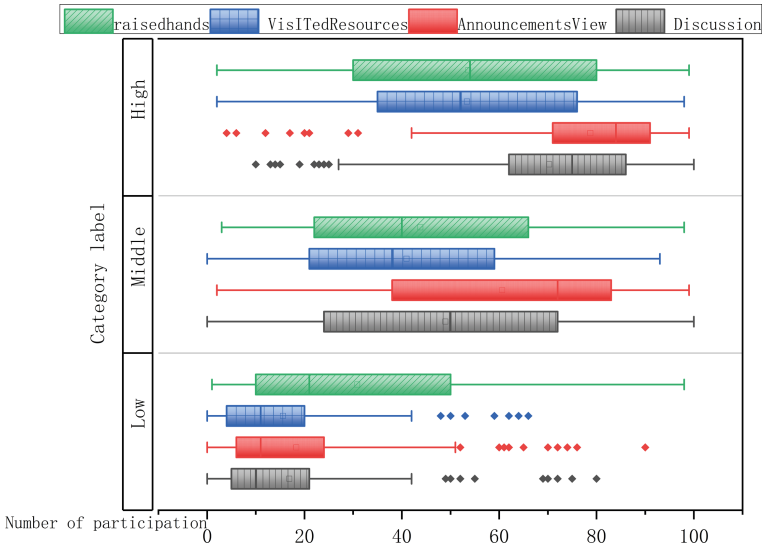


Fig. 1. Box Plot of Numerical Attributes and Students' Grade Levels(owner-draw)

wherein, P_k is the importance degree of the kth input feature; m is the number of input features; n is the number of decision trees; t is the number of nodes of each decision tree; D_{Gkij} is the reduction of the Gini index of the kth input feature at the jth node of the ith decision tree;

3) Find out the least important feature $p = \arg\min_k p_k$;

4) Update the feature set $\mathcal{R}=\{p\} \cup \mathcal{R}$;

5) Remove this feature $\mathcal{F}=\mathcal{F}/p$ from $\mathcal{F}$.

When the feature subset in $\mathcal{R}$ is 5, it has a good performance in prediction accuracy. At this time, the optimal feature subset obtained by RFE_RF is { StudentAbsenceDays, Discussion, AnnouncementsView, VisITedResources, raisedhands}, which is used as the input of the prediction model. Figure 1 can be seen that students' grade interval intervals are strongly and positively correlated with the number of class participation.

3 Weighted Voting Algorithm Based on Information Entropy

For the prediction of students' grade categories, the categories are divided into {low-level interval, mid-level interval, and high-level interval}. Due to the different accuracy of prediction by each base classier, the prediction results may be affected by the base classifier with larger prediction bias if voting is conducted directly. To reduce this error, we, by introducing weighted voting and combining it with information entropy, proposed a weighted voting algorithm based on information entropy (IEWV), selected appropriate weights for each base classifier, and finally compared it with the prediction results with the accuracy of the prediction results of the base classifier as weights. The specific progress of the IEWV algorithm is as follows:

For the three categories of student grade prediction in this paper, when four classifiers are used for voting fusion, the posterior probability matrix of the student samples is calculated according to the classifiers by the following formula (2).

$$\mathcal{P}(x) = \begin{bmatrix} \mathcal{P}_{11}(x)\ \mathcal{P}_{12}(x)\ \mathcal{P}_{13}(x)\ \mathcal{P}_{14}(x) \\ \mathcal{P}_{21}(x)\ \mathcal{P}_{22}(x)\ \mathcal{P}_{23}(x)\ \mathcal{P}_{24}(x) \\ \mathcal{P}_{31}(x)\ \mathcal{P}_{32}(x)\ \mathcal{P}_{33}(x)\ \mathcal{P}_{34}(x) \end{bmatrix}_{3\times 4} \tag{2}$$

For each student sample x, the more dispersed the categorical samples belonging to each category, the less certain the grade prediction level is. The more concentrated the categorical samples belonging to each category, the greater certain the grade prediction level is. For the probability $\mathcal{P}$ in Formula (2), assuming that the classification uncertainty of base learner i for the student sample X is $\mathcal{H}_i(\mathrm{x})$, its information entropy is shown in Formula (3).

$$\mathcal{H}_i(\mathrm{x}) = -\sum_{j=1}^{m} \mathcal{P}_{ij}\log_2\mathcal{P}_{ij},\quad \mathrm{i} = 1, 2, 3, 4 \tag{3}$$

The obtained (x) measures the performance of the ith base classifier. The larger its value, the greater the uncertainty of classification, which indicates that the worse the classification ability of the base classifier for the student sample x is, the smaller the weight of the base classifier should be. The information entropy was converted into the weight of the base learner according to the normalization formula (4).

$$\omega_i = \frac{\exp(-\mathcal{H}_i(\mathrm{x}))}{\sum_{j=1}^{4}\exp(-\mathcal{H}_j(\mathrm{x}))} \tag{4}$$

The new probability matrix $\mathcal{P}'(\mathrm{x})$ can be obtained by multiplying the i-th row of the matrix $\mathcal{P}(\mathrm{x})$ with the corresponding weight ω_i is shown in Formula (5).

$$\mathcal{P}'(x) = \begin{bmatrix} \omega_1\mathcal{P}_{11}(x)\ \omega_1\mathcal{P}_{12}(x)\ \omega_1\mathcal{P}_{13}(x)\ \omega_1\mathcal{P}_{14}(x) \\ \omega_2\mathcal{P}_{21}(x)\ \omega_2\mathcal{P}_{22}(x)\ \omega_2\mathcal{P}_{23}(x)\ \omega_2\mathcal{P}_{24}(x) \\ \omega_3\mathcal{P}_{31}(x)\ \omega_3\mathcal{P}_{32}(x)\ \omega_3\mathcal{P}_{33}(x)\ \omega_3\mathcal{P}_{34}(x) \end{bmatrix}_{3\times 4} \tag{5}$$

The weighted probability can be obtained by summing up each column of $\mathcal{P}'(x)$ is shown in Formula (6).

$$\mathcal{P}'_v(x) = \left[\sum_{i=1}^{4}\omega_i\mathcal{P}_{i1}, \sum_{i=1}^{4}\omega_i\mathcal{P}_{i2}, \sum_{i=1}^{4}\omega_i\mathcal{P}_{i3}\right] \tag{6}$$

Wherein the largest column $\underset{j=1,2,3}{\operatorname{argmax}}\sum_{i=1}^{4}\omega_i\mathcal{P}_{ij}$ in $\mathcal{P}'_v$ is the result of student grade prediction.

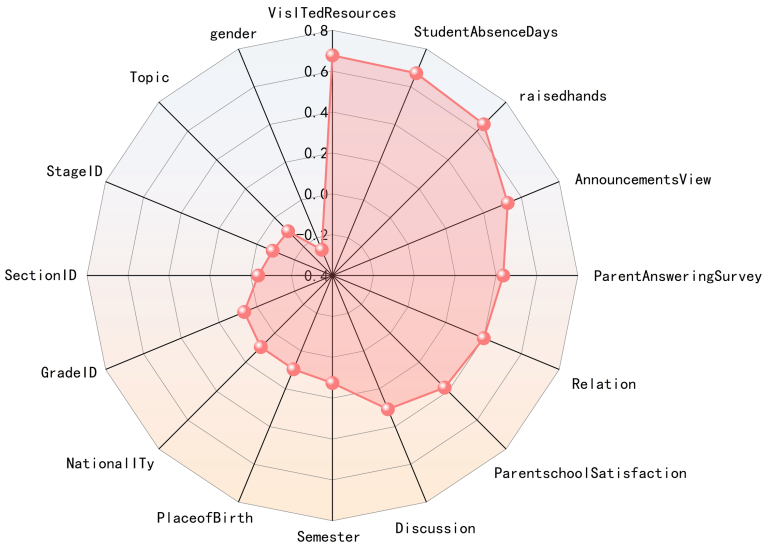


Fig. 2. Correlation Coefficient Plot of Each Attribute and the Interval of Students' Grades(owner-draw)

4 Experiment and Result Analysis

4.1 Data Sources

The dataset (xAPI-Edu-Data) [9] was collected from the Kalboard 360 learning management system through the Learner Activity Tracking Tool. It consists of 245 student records in the first semester and 235 student records in the second semester, totaling 480 student records. According to students' total scores, students were divided into three grades, that is, 0–69 as the low-level interval, 70–89 as the mid-level interval, and 90–100 as the high-level interval. Finally, the correlation analysis of each attribute and the interval of students' grades was carried out, and the resulting correlation coefficient plot is shown in Fig. 2.

4.2 Experimental Design

First, 80% of the dataset was used as the training set and 20% as the test set, and then the training set was divided into a training set and validation set, and a five-fold cross-validation method was used to train the classifier to facilitate parameter tuning. Next, the RFE_RF algorithm was adopted to rank the importance of features and select the attributes that are most effective in improving accuracy. Finally, the weight of each base classifier was calculated by using information entropy, and the predicted value was obtained using the integrated voting algorithm.

4.3 Result Analysis

The test results of the decision tree, random forest, logistic regression, Naive Bayes, voting integration algorithm, and IEWV algorithm are shown in Table 1. Among them,

Table 1. Test Results of Machine Learning

Accuracy	Algorithm					
	Decision Tree	Random Forest	Logistic Regression	Naive Bayes	Voting Integration Algorithm	IEWV
Overall	76.04%	82.29%	80.21%	75%	82.29%	84.38%
Low	93%	88%	90%	88%	90%	93%
Mid	76%	88%	84%	84%	88%	86%
High	62%	72%	63%	59%	70%	73%

logistic regression is the most popular education prediction model at present and is often used as a baseline model [10]. In this sample, the IEWV algorithm model has the best overall accuracy, which correctly identified 84.38% of students' grade levels, and is also the best model for predicting high-level students, with an accuracy of 73%.

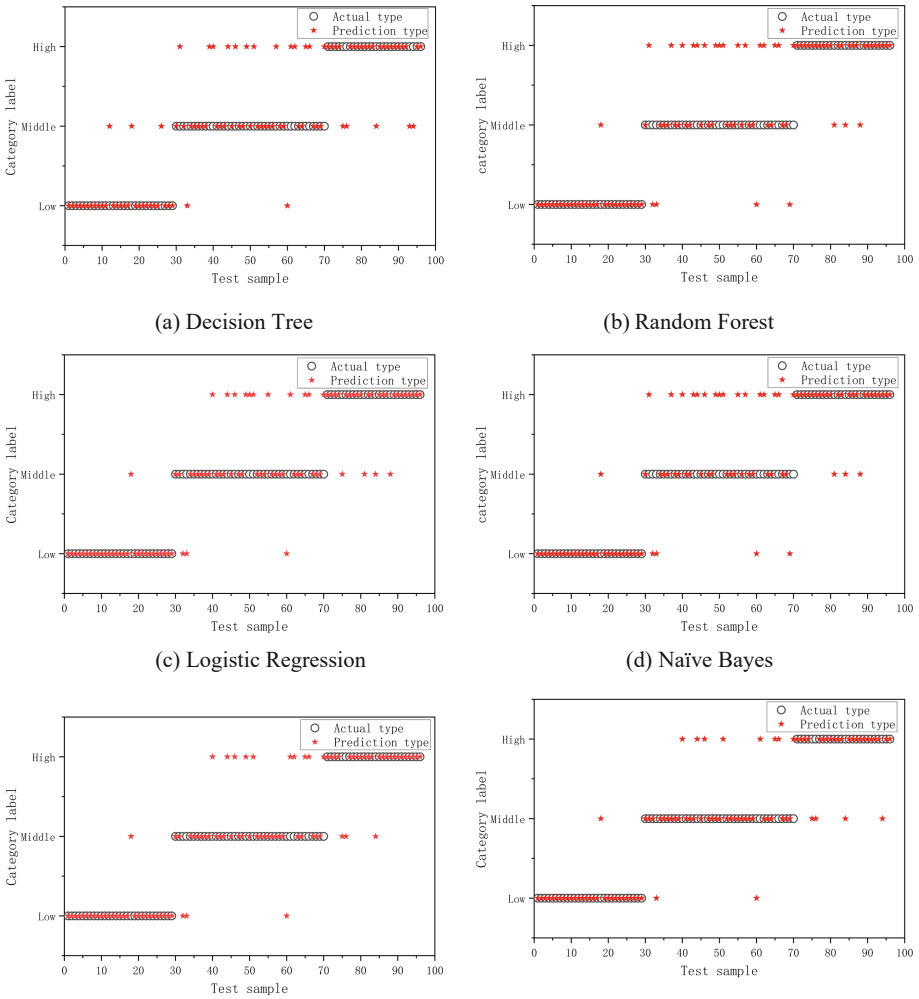
From a theoretical point of view, integrated learning can give full play to the advantages of each learner and reduce the risk of poor generalization performance of a single model to a certain extent. In addition, the prediction accuracy of the high-level interval is quite different from that of the low and mid-level interval, which is mainly due to the generally uneven distribution of data generated in the teaching process and the small number of students with grades in the high-level interval.

Figure 3 provides a comparison chart between predicted grades and actual grades of each model. The acceptable error is considered to be within the errors of high-level students and mid-level students because these students do not have the risk of failing exams and are not among the students under the key observation of the school. However, an unacceptable error occurs between the students in the mid-level interval and the students in the low-level interval. Incorrectly predicting the students in these two intervals may cause some students not to study hard and thus fail exams.

By comparing the number of student samples in the low-level interval that were incorrectly predicted as the middle-level interval for each model, it can be obtained that IEWV only predicted a student sample incorrectly. By comparing the number of student samples in the middle level that were incorrectly predicted as the low-level interval for each model, it can be obtained that the decision tree and IEWV have the minimum error, which predicted only two student samples incorrectly. Combining the two types of prediction errors, IEWV has the best prediction results, followed by a voting integration algorithm and logistic regression.

The recall and F1-score in the IEWV algorithm model prediction results were compared with the prediction indicators of other algorithms, as shown in Fig. 4. It can be seen that the IEWV model ranks first in the recall and F1-score of any student grade level, and performs better than the integrated voting algorithm in all evaluation indicators.

We found that the accuracy of the model was greatly improved by changing the student categories in this study from low, middle, and high to a binary classification model that only predicts whether a student will fail exams, with 0–69 points for failing



(a) Decision Tree (b) Random Forest

(c) Logistic Regression (d) Naïve Bayes

(e) Voting Integration Algorithm (f) IEWV

Fig. 3. Comparison Chart between Prediction and Actual type of Each Model(owner-draw)

exams and 70–100 points for not failing exams. The confusion matrix of the prediction results is shown in Fig. 5. Among the samples with prediction results of failing exams, only one sample was wrongly predicted; among the samples with prediction results of not failing exams, three samples were predicted incorrectly. The accuracy of predicting students who will not fail is as high as 98%, and the overall accuracy of predicting students' grades is 96%.

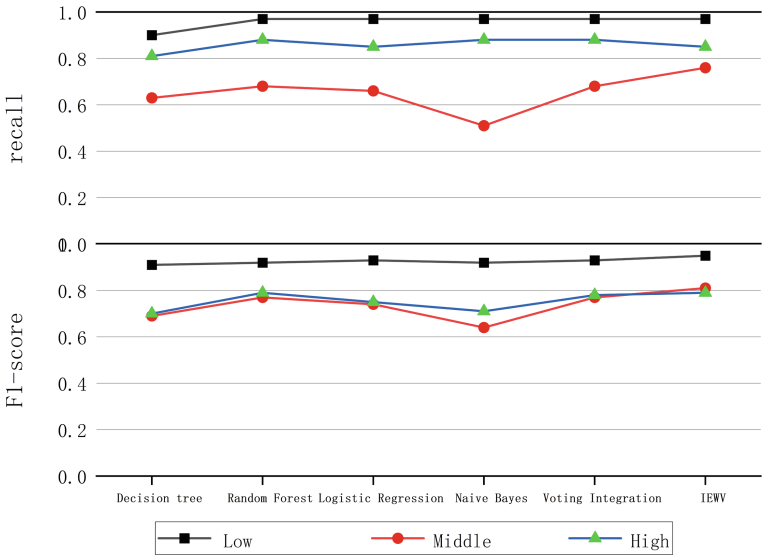


Fig. 4. Comparison Chart of Recall and F1-score(owner-draw)

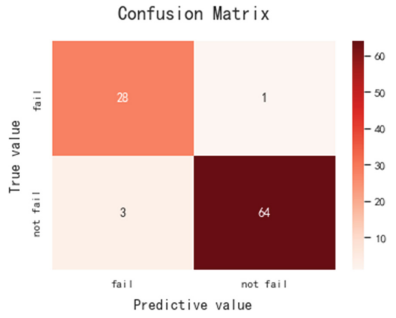


Fig. 5. Prediction Result of Binary Classification(owner-draw)

5 Conclusion

Through analysis, the prediction model based on RFE_RF and the weighted voting algorithm based on information entropy show improved performance to some extent compared with the prediction results of other single models. For the dataset used in this paper, the number of days students missed class, the number of times they participated in class discussions, the number of times they answered questions in class, the number of times they checked course materials, and the number of times they checked notices have the greatest impact on academic performance, followed by whether parents are satisfied with the school and whether parents could complete the questionnaire provided by the school, while students' demographic characteristics have minimal impact on academic performance. This finding has reference significance for online education.

References

1. Li Xue,Teng Da. Construction of Intelligent Education Service Ecosystem in Higher Vocational Colleges in the Information 2.0 Era[J].Education and Vocation,2022(11):28–34.DOI:https://doi.org/10.13615/j.cnki.1004-3985.2022.11.003.
2. Tang Lingqiu. Research on Provincial Education Big Data Platform Construction under the Background of Education Informatization 2.0[J]. Forum on Contemporary Education.,2021(04):99–106.DOI:https://doi.org/10.13694/j.cnki.ddjylt.20210525.001.
3. LU Genshu. Applications and Challenges of Big Data in Higher Education[J]. Chongqing Higher Education Research.,2022,10(04):31-38.DOI:https://doi.org/10.15998/j.cnki.issn1673-8012.2022.04.004.
4. Gonzalez-Nucamendi Andres,Noguez Julieta,Neri Luis,Robledo-Rella Víctor,García-Castelán Rosa María Guadalupe,Escobar-Castillejos David. The prediction of academic performance using engineering student's profiles[J]. Computers and Electrical Engineering,2021,93.
5. Dong D , Li J , Wang H , et al. Student Behavior Clustering Method Based on Campus Big Data[C]// 2017 13th International Conference on Computational Intelligence and Security (CIS). IEEE Computer Society, 2017.
6. SHEN Hang-jie;JU Sheng-gen;SUN Jie-ping. Performance prediction based on fuzzy clustering and support vector regression[J]. Journal of East China Normal University(Natural Science).,2019(05):66–73+84.
7. Jakub Kuzilek, Zdenek Zdrahal, Viktor Fuglik,Student success prediction using student exam behaviour,Future Generation Computer Systems,Volume 125,2021,Pages 661–671,ISSN 0167–739X.
8. Mubarak Ahmed A.,Cao Han,Hezam Ibrahim M.. Deep analytic model for student dropout prediction in massive open online courses[J]. Computers and Electrical Engineering,2021,93.
9. Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), 119-136.
10. Farshid Marbouti,Heidi A. Diefes-Dux,Krishna Madhavan. Models for early prediction of at-risk students in a course using standards-based grading[J]. Computers & Education,2016,103.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.