

Analysis of School District Fiscal Data and its Implication on Student Graduation Rate

IST718 Project - Professor: Daniel Acuña, Presented by Chun Xie, Xuehan Chen, Yimin Xiao, Yue Wang

■ Introduction and research background

This project is an investigation of the relationship between school district fiscal data and graduation rate. It aims to : 1. predict student graduation rate based on fiscal data, 2. infer how fiscal factors affect graduation rate, and 3. cluster school districts to help with policy and decision making.

■ Data and preliminary analysis

Fiscal and graduation data for school districts in the year 2007 to 2010 was both collected from National Center for Education Statistics [1][2]. Figure 1 is a description of the data items and figure 2 and 3 illustrate the distribution of school revenue resources and budget allocations calculated in percentage.

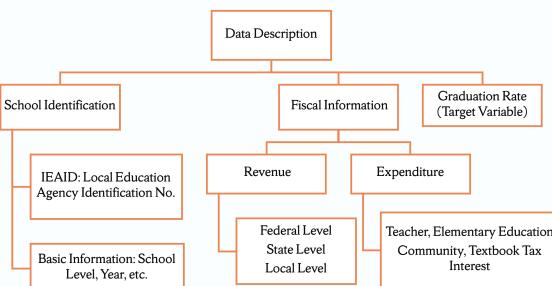


Figure 1 Description of data items and relationship

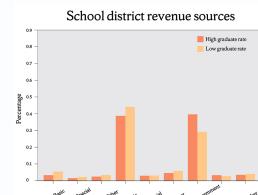


Figure 2 School district revenue sources distribution

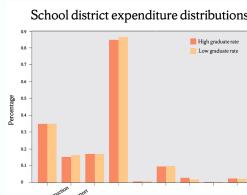


Figure 3 School district expenditure distribution

■ Model description and evaluation metrics

The following table gives an overview of the models built. Classification models used 0.6, 0.3, 0.1 splitted data and clustering model was fitted on all data.

Table 1 Description of classification and cluster models built

Model	Objective	Features	Metric	Performance
Classification-Logistic Regression	Inference Prediction	School district information, all fiscal data	AUC	Validation: 0.764 Test: 0.757
Classification-Random Forest	Prediction Inference			Validation: 0.801 Test: 0.801
Classification-Gradient-Boosted Trees	Prediction Inference			Validation: 0.800 Test: 0.800
Clustering - K-Means	Inference		None	None

■ Model 1: Logistic regression

The logistic regression model was firstly fitted for both inference and prediction. Figure 4 and 5 contain a list of most positive and negative features from the model.

features	weight
8 Re_L_fee_Per	13.477682
10 Ex_Teacher_Inst_Per	1.832259
14 Ex_Community_Per	1.370589
13 Ex_Edu_Per	1.150229
25 AGCHRT2	0.837651

Figure 4 Top 5 positive features - LR

features	weight
17 Ex_Textbook_Per	-26.100121
1 Re_F_Basic_Per	-25.849196
11 Ex_Teacher_Supp_Per	-7.136946
3 Re_F_Other_Per	-6.375287
6 Re_S_Other_Per	-2.586610

Figure 5 Top 5 negative features - LR

■ Model 2: Random forest

The random forest model was then fitted to build a more accurate classification model. Figure 6 is a list of the five most important features in the bst model generated.

features	weight
1 Re_F_Basic_Per	0.209695
0 V33	0.100442
7 Re_L_Gov_Per	0.083649

Figure 6 Top important features - RF

features	weight
1 Re_F_Basic_Per	0.112664
0 V33	0.101049
5 Re_S_Special_Per	0.071157

Figure 7 Top important features - GBT

■ Model 3: Gradient-boosted trees

The GBT model was also fitted to improve prediction accuracy. Figure 7 is a list of the five most important features in the bst model generated.

■ Model 4: K-Means clustering

The K-means clustering model uses data for school districts in New York State in 2010, to explore if districts of different poverty level exhibit different fiscal plans. Following figures show the cluster result for all districts and those with low, medium, and high poverty level [3].

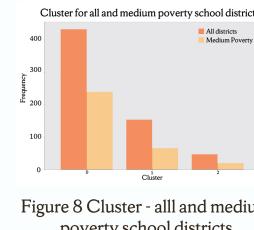


Figure 8 Cluster - all and medium poverty school districts

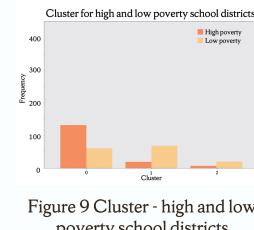


Figure 9 Cluster - high and low poverty school districts

■ Summary and conclusion

Our analysis has produced three classification models with test AUC of 0.757, 0.801, and 0.800 respectively, and a clustering model. The models also provide an inference on the relationship between fiscal data and graduation rate.



Figure 10 Inference on data analysis result

■ References

- [1] 'Local Education Agency (School District) Finance Survey (F-33) Data'. National Center for Education Statistics. Retrieved online on Feb 15th via <https://nces.ed.gov/ccd/f33agency.asp>
- [2] 'Local Education Agency (School District) Universe Survey Dropout and Completion data'. National Center for Education Statistics. Retrieved online on Feb 15th via <https://nces.ed.gov/ccd/drpagency.asp>
- [3] 'New York School District Demographic Characteristics'. ProximityOne. Retrieved online on Apr 12th via http://proximityone.com/sd_ny.htm