Comparison of the Outcome of Different Machine Learning Classifiers: Using the Datasets on School Funding Requests

Introduction.

Our client, which is a school district administration, is interested in the question that what types of schools, as well as what kinds of funding requests, are likely to get the funding within 60 days. They provide us with a full dataset on school funding requests, which consists of information on school name, school location, requested resource type, the poverty level of the area where the school is in, request date, date when the school gets funding, etc.

Our approach.

The data is from January 1st, 2012 to Dec 31st, 2013. As it is a time-series data, we decide to use temporal validation which creates training and testing datasets over time. We decide each testing set to be six months long, and the training set is all the data prior to the beginning of each testing set. As a result, we have three training-testing pairs. For the machine learning models, we include bagging classifier, random forest classifier, logistic regression classifier, support vector machine, gradient boosting classifier, decision tree classifier, and k-nearest neighbor classifier. Moreover, in order to guarantee the objectivity of our studies, we also use different parameters for each model we use; for instance, in logistic regression classifier models, the regularization methods include Lasso Regression (or 'L1' regularization) and Ridge Regression (or 'L2' regularization), so we tried both regularization parameters in our studies. We choose 'poverty_level' and 'total_price_including_optional_support' to be the features of our models, as we assume that poorer schools who apply for funding of the projects with higher costs are not likely to get their funding fulfilled in 60 days.

Findings.

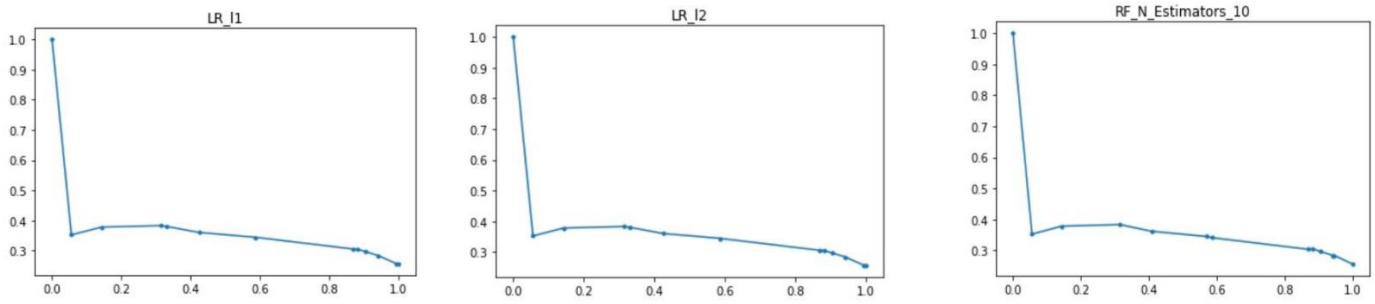
In our first test (training set: data from 2012/01/01 to 2012/06/30; testing set: data from 2012/07/01 to 2012/12/31), our accuracy scores range from 0.7343 to 0.7430; however, most of our F1 scores are

zero. The situations are similar in the second test and third test. That is to say, the results do not vary significantly over time.

The result of the first test:

	model_type	accuracy	f1_score	auc-roc	p_at_5	p_at_10	p_at_20	r_at_5	r_at_10	r_at_20
0	BG_N_Estimators_10	0.743072	0.000000	0.636786	0.318708	0.368261	0.347190	0.061989	0.143297	0.270238
1	BG_N_Estimators_100	0.743072	0.000000	0.636252	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
2	RF_N_Estimators_10	0.743072	0.000000	0.636252	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
3	RF_N_Estimators_100	0.743072	0.000000	0.636786	0.318708	0.368261	0.347190	0.061989	0.143297	0.270238
4	LR_l1	0.731043	0.096748	0.637947	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
5	LR_l2	0.731043	0.096748	0.637947	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
6	SVM_C_l1	0.743072	0.000000	0.637947	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
7	SVM_C_l2	0.743072	0.000000	0.637947	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
8	GB_N_Estimators_10	0.743072	0.000000	0.636786	0.318708	0.368261	0.347190	0.061989	0.143297	0.270238
9	GB_N_Estimators_100	0.743072	0.000000	0.636252	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
10	DT_GINI	0.743072	0.000000	0.636252	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238
11	DT_ENTROPY	0.743072	0.000000	0.636252	0.288239	0.368261	0.347190	0.056063	0.143297	0.270238

The precision-recall curves of the Logistic Regression model and Random Forest model in the first test:



Conclusion.

To conclude, none of the models we use perform idealistically in this machine learning studies, no matter what evaluation metrics that we refer to. In the first test, even if the accuracy scores are relatively high, which is approximately 0.74, it does not necessarily indicate that our models do a good job in prediction. If you look at the precisions and recalls at 5% level, all of them are below 0.5. It means that the possibility that the model successfully predicts if a school receives the funding that they want within 60 days is below 50%, which does not meet our client goals of "identifying 5% of the posted projects to intervene with". The results tell us that our assumption might not reflect the real-life scenarios.

Recommendations.

In order to build better machine learning models, we recommend our client to provide us with raw data that contains more useful information, such as the teacher's qualifications, the school's ratings within

its school district, the donor's wealth, etc. More information would be useful for our analysis. Also, as for the models that we could go forward with and deploy, we would not suggest Logistic Regression Classifier. If we incorporate more features into our machine learning models, the decision boundaries might not be linear; the Logistic Regression might underperform in such situations. The models that we would recommend include Random Forest Classifier, Bagging, and K-Nearest-Neighbors, because they are not only performing slightly better in our studies, but also robust to large and noisy training data.