Competition 04: Google Analytics Customer Revenue Prediction

Team: C4-11 Sakshi Jawarani, Bowei Sun, Jingyi Luo

Who might care about this problem and why?

Google Merchandise Store cares about the prediction of revenue for each customer. The results will help the company better prepared marketing budgets that may attract more customers and boost the overall revenue. The model may also be used to guide future decision making regarding the products to attract customers produce most of the revenue. Identifying customers who are not so interested in the products would also help the company improve its products for these customers.

Why was this problem challenging?

The data contain different types and there are few or no analytical model that can full advantage of all the data available. The exact relationship between the revenue and the variables are not clear. The data have many missing values. It is unclear which variables are more useful in predicting the revenue and which are not so useful.

What other problems resemble this problem?

Predicting revenue for online shopping/ordering websites resemble this problem. For these websites, it will be useful to how which group of customers buy more products so that the companies could customized the product for these customers. As for the not-so-interested customers, the companies may want to identify them as well. Different types of products could be offered to these less-interested customers.

What might account for the differing performance levels of the mandatory models?

First, the linear model and nonlinear model performance differently. As the relationship between the revenue and the variables are not known, the linear model and nonlinear models have different performance level. Second, how the model is constructed also matters. For example, Random Forest works well for multiclass problem while SVM are usually used for two classes.