ISYE 6501 Homework 12

Question 18.1

Describe analytics models and data that could be used to make good recommendations to the power company. Here are some questions to consider:

• The bottom-line question is which shutoffs should be done each month, given the capacity constraints. One consideration is that some of the capacity – the workers' time – is taken up by travel, so maybe the shutoffs can be scheduled in a way that increases the number of them that can be done.

• Not every shutoff is equal. Some shutoffs shouldn't be done at all, because if the power is left on, those people are likely to pay the bill eventually. How can you identify which shutoffs should or shouldn't be done? And among the ones to shut off, how should they be prioritized?

Think about the problem and your approach. Then talk about it with other learners and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}. Have fun! Taking a real problem and thinking through the modeling and data process to build a good solution framework, is my favorite part of analytics.

Solution

Step One:

Given customers data such as:

- Payment history
- Outstanding balance on the account
- Credit score
- Income
- Default on other accounts/bills
- Employment status
- How long the customer has been with the company

We can **use** a clustering model,

To divide these customers in 3 groups based on the possibility of not clearing the outstanding dues: low, medium and high.

Results: After obtaining these clusters, we can carefully examine the customers in each group. For people in low risk category, we can set some strategies such as following up with customers regularly to

remind them of the payment. Also, we should make sure all information which we are using in our clustering model is up to date.

For people in medium risk category, we can also do follow up. However, we can use this subset of data as input for classifying customers to find out which ones will never pay off their debt.

We can do a similar model for people in high risk category and categorize them as will pay/will never pay.

Although, we could use mutli-category classification too, to find three classes of low, medium and high-risk clients, the advantage of clustering is that it might reveal some unknown/hidden patterns since it is unsupervised.

Step Two:

Given subsets of medium-risk and high-risk customers (2 clusters obtained from previous model),

We can **use** support vector machine (SVM) classification,

To classify customers as will/will not pay.

The independent variables could be income, credit score, amount of debt, default on other accounts and history/pattern of power usage and how long that person has been a customer.

Results: All the customers classified as people who will never pay their debt, can be collected together and form a data set for further cost analysis. We can continue with building an optimization model to find customers who need to be listed as shutoff ones.

Step Three:

Given customers classified as “will never pay” from previous model,

We can **use** exponential time series,

To predict the coming months power usage by each customer.

Results: The predicted power usage by customers who are not going to pay for it will be used in an optimization model along with other associated costs to find out how the utility can maximize the profit while minimizing the cost of shutoff.

Step Four:

Given those customers predicted power usage, outstanding balance on the account, cost of assigning workers, logistical problems, and all other expenses,

We can **use** a logistic regression,

To classify customers whose shutoff cost is less than cost of keeping the power on.

Results: The profit or saving that the utility will gain can be defined as a new variable which is the difference of cost of shutoff and cost of keeping the power on. This variable will be included to our data set. Based on how much saving will justify the decision to disconnect the power, the logistic regression can classify a final list of customers whose power should be cut off.

Finally, a discrete-event simulation can be performed to find the optimal number of workers and vehicle routing for scheduling the shutoff.