Power Company Case Study

Design Approach

As a recap, the goal of the power company to initiate this program is to minimize the loss of profit from supplying the power to the defaulters who are financially capable of paying the bill but didn't pay.

Given that, the solution could be separated into 3 steps:

1. Calculate the probability of default of each householder.
2. Classify the potential defaulters into different groups and identify the "target" defaulters.
3. Optimize the shutoffs selection to maximize the expected value. (Or Minimize the expected cost)

(*In the following demonstration, I will use 'GIVEN- USE- TO' format to better illustrate my thoughts)*

First, step 1: Calculate the probability of default of each householder.

Given{ income, credit score, bill payments history data, age, marital status, number of people living in the same household, rent(0) or own(1) of the house}
Use(Logistic Regression)
To(estimate the probability of default)

In this model, we need to make variable selection along the way by ruling out the non-significant factors among the initial inputted factors. Several iterations may be suggested.
Also, since it's a supervised learning method, the response of the logistic regression is 1 (If they default in the most recent month) or 0 (If not).

After coming up with certain probabilities for each household, a threshold could also be determined by highest AUC.

Second, step 2: Classify the potential defaulters into different groups and identify the "target" defaulters.
Remember, from the above model, the output (probability > threshold) is the potential "defaulters", not "abusers", since it includes the households that are not financially capable of clearing the bill each month, the households that are just missed the due date and will pay their bills soon and the possible "abusers" . Therefore, I need to classify the output(household) from above model into 3 categories and identify the potential "abusers".

Categories listed as below:

1. Financially struggled defaulters

2. Due-date missing defaulters
3. Abusers

Given{ probabilities of default > threshold, income, credit score, bill payments history data, age, number of days of defaulting until bill is cleared during last default, monthly power usage. }
Use(KNN)
To(Take the output from logistic regression and classify the defaulters into 3 groups)

Finally, step 3: Optimize the shutoffs selection to maximize the expected value. (Or Minimize the expected cost).

Due to the constraints of the work force capacity, it's not worthy to shut off the power of each "abusers". Several factors came into play when deciding the shutoff schedule.

1. We first need to make a prediction of the power usage amount for these abusers in the following month to estimate our potential loss if their power is not shuttled off.

Given{ probabilities of default > threshold, income, bill payments history data, age, number of days of defaulting until bill is cleared during last default, monthly power usage. }
Use(Exponential smoothing)
To(estimate the amount of power a customer will use in the next month)

2. Then, the cost of sending off a technician to certain place to shut off their power could be modelled using regression model.

Given{distance between power company and customers' location, number of days for a round trip, number of potential abusers in the near city(<200KM), travel expense per trip, technician's salary }
Use(simple regression)
To(estimate the labor cost of shutting down the customers' power)

3. Finally, use build an optimization model.

Constants:
m= estimated cost of power usage in the next month
c = estimated labor cost of shutting down the customers' power

Variables: x= number of company technicians to shut down customers' power
A_j= 1, if shut down customer j's power, 0 if not
Constraints: x >=0
Objective function: maximize the potential gain by varying the number of employees to send off.