**Homework Week 8**

**Question 18.1:**

Here are the steps I would recommend to the power company:

1. **Step One: Calculate the Probability of Not Paying the Bill on the Date when the Model is Run**

**Data:**

1. For each customer who hasn’t paid their bill for current month, we need the number of days between their bill payment date and bill due date for every month in history. For example, if in the last month, the customer paid the bill one day before the due date, then the data we collect for this payment would be -1. If the customer paid the bill three day later than the due date, then the data we collect should be 3.
2. For each customer, we need to calculate the number of days between the date when the model is run and their current bill due date.

**Model:** Probability distribution.

**Results / Target:** For each customer, predict the probability that they are not going to pay their current bill when the model is run.

**Description:**

We are going to use a similar approach to the “seat upgrade” problem in lecture “Module 13 – Part 1”.

We are going to fit the first (1) data we collect for each customer to probability distributions. Then we are going to use the second (2) data we calculate, along with the probability distributions we derived, to estimate the probability that a customer is not going to pay their current bill when the model is run.

The reason why I choose this approach instead of any other more complicated approach like logistic regression is because the data used in this approach is easier to collect.

**Additional Concerns:**

Some new customers may not have enough payment history for us to use this approach. In this case, we can use “K nearest neighbor” model to estimate the probability distribution for the new customer using the data of similar existing customers. Some additional customer info may be needed as data input of the “K nearest neighbor” model, such as customer’s zip code, number of people in the household, household income, etc.

1. **Step Two: Prioritize the Shutoffs**

**Data:**

1. The probability of not paying the bill, which has been calculated for each customer in step one.
2. The billed amount for each customer in the current month.

**Results / Target:** Create a new indicator, Expected Unpaid Amount, to help us prioritize the shutoffs.

**Description:**

We need to define a new indicator, and use it to determine how to prioritize the shutoffs:

When the power company send out their workers to cut off the power, they will make sure that the customers with higher Expected Unpaid Amount will be prioritized to cut off.

1. **Step Three: Determine which Customer’s Power should be Cut Off**

**Data:**

1. The probability of not paying the bill, which has been calculated for each customer in step one.
2. Locations of customers (Locations to be shut off. Only customers whose due date has passed when the model is run will be added to this dataset.)
3. Locations of workers
4. The time to travel between each pair of locations (including customers’ locations and workers’ locations)
5. Expected Unpaid Amount for each customer we calculated in step two.

**Model:** Network Optimization

**Results / Target:** Determine the customers that should be shut off by maximizing the total Expected Unpaid Amount of customers who are successfully shut off while meeting serval constraints.

**Description:**

We are going to use network optimization algorithm to resolve this issue. Here are more details about this algorithm.

**Variables:** The route that each worker will take to travel between different locations.

**Target:** Maximize the total Expected Unpaid Amount of customers who are successfully shut off while meeting serval constraints.

(This approach will not guarantee that customers with higher Expected Unpaid Amount are shut off before lower ones. It will also not guarantee that the number of shutoffs is maximized, but it will generate an overall optimal solution for the entire power company.)

**Constraints:**

1. Each worker will eventually travel back to the original location where they started.
2. Each worker should not spend more time on travel than a budget number of minutes.
3. The power company can set up a threshold for the probability of not paying the bill. For example, they can let the workers shut off the power of customers who are at least 70% likely not to pay their bills. The power of the other customers will not be shut off.

**Additional Concerns:**

Some other input can also be taken into consideration. For example, if a customer has never been cut off due to not paying the bill before, then the power company may not cut off their power immediately even if the network model indicates that this customer should be shut off. The power company can choose to send a notification to the customer and give them a “grace period” of one week.