Intro to Analytics Modeling HW 8

Power Company Case Study

2024-07-09

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.

Breakdown of the Problem Statement

1. The company wants to turn power off for those who are able to pay but are not ever going to pay, but the company wants to keep the power on the the customers who might have just forgotten or get behind but are able to pay it all back after a while

2. Prioritize the customers whose energy will be shutoff

3. The shut-off had to be done manually by workers who had to go to each location

    1. The time required to travel all around the city and turn off the power at all those locations was more work than the company had the capacity to do because there are more people each month whose power should be shut off than the company had capacity to handle

Proposed Solution

Identify which shutoffs should be done or shouldn't be done

- Required data (for each customer)
    - Income-based rate classes: can be used to identify if the customer is eligible for the assistance programs so that they don't need to pay the bills
    - Monthly payment history (On-time payments, energy usage, total bill amount): gives an idea of the payment habit whether the customer has been responsible for the payment or not.

Given the set of required data, we can first filter the customers using income-based rate classes to keep only those who are able to pay the bill. Additionally, by using monthly payment history data, we can cluster our customers based on their payment behavior. These clusters should indicate how responsible the customers are when it comes to their bill payments historically.

To achieve this, we can transform the monthly payment history data into metrics such as consecutive months overdue, the number of on-time payments in the last year, and the number of missed payments in the last year. We can then use the k-Means clustering algorithm to group the customers into clusters. There should be at least one or two clusters of customers whose energy the company can proceed to shut off.

We have identified a group of the customers to shutoff. Let's prioritize them now.

Now that the company's energy capacity is the main issue for this shutoff activity, monthly energy usage for each customer matters the most.

- Required data (for each customer)
    - Monthly energy usage

Given the data, we can use the ARIMA model to predict next month's energy usage for each customer. This should be applied only to customers whose energy will be shut off. Once we have the predicted energy usage for each customer, we can sort them in descending order to prioritize those with higher energy usage.

Let's schedule workers time and travel

We need to schedule workers time and travel in a way that increases the number of shutoffs that can be done.

- Required data (for each customer)
    - Travel time matrix between customers

Given the travel time matrix data, we can use the Louvain algorithm to identify communities and the total travel time within each community. What we can do with the outcomes of the algorithm are as follows:

1. Compare the number of the communities and the number of workers. If the number of workers is less than the number of communities, we can combine neighboring communities together to match the numbers. Or, if the number of workers is greater than the number of communities, we can break big communities into a smaller communities by running the Louvain algorithm again for the big communities. Either of these two should be done until the numbers match.
2. Recalculate the total travel time within each community.
3. For each community, we need to find the fastest route for each worker based on the importance of the customer that can be defined by the predicted energy usage next month.
    1. In order to find the fastest rout based on the travel time and the importance, we need the following set of data
        1. Travel time matrix between customers
        2. Predicted energy usage
    2. Given the set of data, we can use the Dijkstra's Algorithm to find the fastest route.