Homework week 8 **July, 8 2020**

Valerio Berardi

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.

Answer.

I have divided my solution in 5 steps:

1. Identify customers that don't pay and are suitable for shutoff
2. Calculate the monthly use of power of those customers
3. Clustering of customers who don't pay
4. Optimize/Prioritize the shutoff
5. Run a simulation

Step 1: Find defaulting customers

Given..	Use..	To…
Customer Information	K-nearest neighbors (or Classification Tree/Random Forest)	Classify customers into 3 categories: • Customers who regularly pay • Customers who are never going to pay • Customers who don't pay for god reason/ will pay later

Customers information:

- Credit history
- household income
- Previous delays in payment history (binary)
- Number of months of current defaulting payment
- Customers' eligibility for Low Income Home Energy Assistance Program (binary variable)

- Customers' loyalty (length of account with the power company)
- Type of service linked to the address (e.g. residential, business)
- Ownership of the property (binary variable)

Collecting personal information such as race, sex and other correlated demographic factors is illegal in many countries, therefore I chose predictors the power company would most likely already have or could easily collect. KNN method is sensitive to outliers, so preliminary analysis of the data will be required. In the case of missing data, I believe the best approach should be the imputation of an estimated value (regression). The company should consider shutoff for customers in the second category.

Note: KNN is a method for hard classification. To be more flexible in the response, an alternative is the use of a method with an output in terms of probabilities, like a Classification Tree or a Random Forest Model. Using this second approach, it is possible to modulate the response based on the company budget and request by changing the threshold in the response.

Step 2: Estimation of cost

Given..	Use..	To…
Customer Information	ARIMA	Estimate power usage per account each month

Additional customer information:

- Past power usage
- Season

After customers are identified as possible target for power shutoff, it is possible to use ARIMA to predict the amount of power they will use each month. The power company should consider to cut off the power to defaulting customers with high power use/month first.

The results will be used in the optimization step.

Step 3: Clustering

Given..	Use..	To…
Results of steps 1, 2 and collected info	K-nearest neighbors	Geographically cluster the customer

Additional information to consider:

- customers address
- drive time/cost
- workers

In this step the customers already identified as non-payers (step 1) are divided in clusters according to the geographic distance between customers, the cost and speed of shutoffs (e.g. drive time, and cost of workers), the estimation of power usage of the next month (step 2). In case the step one gives an output in terms of probabilities, we can use different thresholds to design the clusters in terms of classification probability and cost. The number of clusters can be

modulated based on the capacity constraints of the company. The power company should consider to prioritize the shutoff not just by the power usage of the customer, but also in terms of "accessibility" of the customers for the workers that are performing the job. In this sense, the shutoff should be done inside the clusters, and the priority order should be for cluster and not for customer.

Step 4: Optimization to find the priority shutoff

The final goal of the company is to minimize the cost of the shutoff operation, and this topic can be solved with an optimization model.

Given..	Use..	To…
Results of steps 1, 2, 3, customers info company info	optimization	Minimize the cost of the shutoffs

Variables:

- w= number of workers needed
- x_i= shutoff of customer i

Constrains:

- each worker cannot drive more than 100 Km/day
- workers contract restrictions (e.g. person can work 5 days a week and 8 hours a day)
- number of workers ≥ 0

Objective function:

Min $\sum c_i x_i$

With: c_i = shutoff's cost of customer i

Additional factors to consider:

- cost of driving (cost per Km)
- cost of workers (e.g. $ per hour)
- number of workers available
- time necessary to cut the power
- cost of turning off the power

In this step it is found which shutoff should be performed and the number of workers needed to run the task. The optimization can be run in terms of clusters: we can use the optimization to find the optimal order of priority for the clusters (which cluster should be handled first), and we can use an optimization model to determine the optimal way to perform the shut-down inside a specific cluster.

Step 5: simulation

As a final step, I believe it is good to run a simulation to analyze the performance of this approach. I can use a stochastic simulation to include randomness into the system. The simulation helps to find the more effective way to perform the shutoffs, for example, by finding the fastest route to shutoff the power (the faster route is better than the shorter, because it is more cost effective), or it can help to manage the workflow.