Week 8 homework

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.

Student Answer

First, we determine which customers we should perform shutoffs for, as well as how we will prioritize customers for shutoffs.

Given*:*

- Previous payment data (payment amount, number of missed payments, time between billing and payment)
- Power usage
- Customer demographic data (age, marital status, education)
- Property type (residential, commercial, industrial, etc.)
- Ownership type (rented vs owned)
- If residential, number of residents

Use: A classification model, such as SVMs, logistic regression, or a decision tree. In fact, try multiple classification models before validating and selecting the final model.

To: Predict the probability that customers will default on payment.

Notably, we will not consider customers who have paid all their bills thus far. These customers are not relevant to our model. Instead, we want to separate the remaining customers into those that *will eventually pay* and those that *will no longer pay.*

This model can be applied to new customers to predict their likelihood of paying their bill, or if existing customers are beginning to show signs that they will stop paying their bill. By predicting and identifying defaulters early, the company can determine customers to perform shutoffs for before they incur higher losses.

Also, we can use the predicted probabilities to calculate the expected losses for the company. This will also help us prioritize the customers to perform shutoffs for. For example, consider two customers with similar probabilities of defaulting: Customer A who is 75% likely to miss the next payment and Customer B who is 70% likely to miss the next payment. However Customer A has a low electricity consumption of 800kWh per month while Customer B has a much higher consumption of 3000kWh and has incurred a high bill. In such a case, although Customer A has a higher probability of defaulting, Customer B will result in higher losses for the company, and therefore should be prioritized for shutoffs over Customer A.

After identifying customers for shutoffs, we can then start to find the optimal solution for distributing resources to perform shutoffs.

Given:

- Location data of the property (latitude and longitude)
- Time to travel between property locations
- Location data of employees (eg. where each employee is based within a city)
- Company resources for manpower and transportation

Use: Clustering models: K-Means algorithm, Network optimization program

To: Generate clusters of property locations, then determine the optimal way to distribute manpower resources between these clusters.

First, we will want to use a clustering model to determine clusters of property locations that we have to send employees to. The benefit of using a k-means clustering algorithm instead of a more advanced algorithm such as the Louvain algorithm is that we are able to set the desired number of clusters for the k-means algorithm. This way, we will not be stuck in a situation where we have too many or too few clusters without the ability to adjust.

Once we have our clusters, we can get data on the time it takes to travel between locations within each clusters. This can be done using publicly available data such as Google Maps or GPS, or can be calculated in-house if preferred.

Then, we can implement a network optimization program for distributing resources. Some variables include the travel time to, and between, locations, the fuel cost to travel, and the cost of hiring one more employee. We will attempt to maximise the number of shutoffs we can perform while implementing constraints on time and monetary cost.

If we are unable to find a satisfactory solution, we can go back to the k-means algorithm to change the number of desired clusters before running the optimization program again.

Conclusion:

A combination of a classification model, a clustering algorithm and a network optimization program will help us determine which shutoffs should be performed for the month.

We can use a classification model to estimate which customers will cost the company the most money, which will be used as the basis for prioritizing shutoffs.

We also used a clustering algorithm to identify clusters of property locations. The selected clustering algorithm in this case was the k-means algorithm for it's suitability with geolocation data and the control we have over the number of clusters.

After identifying clusters, we use a network optimization program to decide how many employees we will need and where to send them.

Although not included in the solution, we could also take into consideration variations in travelling time based on weather or seasons as travelling time and convenience during some weather events may take longer or may not be possible at all.

Something else to consider would also be the ethical implications of shutting off a customer's electricity. While we deal with a lot of hard data in analytics, there are some things that are difficult to quantify and it's important to consider non-numerical factors that may not be included. For example, instead of initiating a shutoff at the first offence, an effort could be made to extend help and resources to customers that may not know they exist. While there is a fund set up for customers who cannot afford to pay, some customers that cannot afford to pay may not qualify for the fund, or may not know it exists. It's also important to consider when the shutoffs are being performed. Shutting off power during the winter months can be dangerous for residents of a building if it leaves them without access to heating. Thus I believe it may be prudent to exercise a small amount of leniency when considering which shutoffs to perform that we may not be able to account for through analytics.