

# GT Introduction to Analytics Modeling - Week 8 HW

## Introduction

For this homework, we will propose analytics models and data that could be used to make good recommendations to the power company with regards to shutoffs. Items to consider are:

1. 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.
2. 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?

## Solution Summary

A summary of the models in this solution in the {given, use, to} format are as follows. Note that more details are provided in the next section.

### Model 1 - Payers and Non-Payers

#### Given

- Historical Power Consumption per month per residence (address)
- Payment History of current resident (indication of paid, paid late, or not paid)
- Resident status (rent / own, # of months at residence, # of occupants)

#### use

- A logistic regression model

#### to

- Determine a probability for payers / non-payers

### Model 2 - Average Payment Amount

#### Given

- Historical Power Consumption per month per residence (address)

#### use

- An averaging model

#### to

- Estimate a payment amount for subsequent months

### Model 3 - Worker Shutoff Utilization

#### Given

- Estimated payment amount
- Power company employee information (count of employees, working hours)

#### use

- An optimization model

to

- Determine areas workers should perform shutoffs

## **Solution Description**

This section provides a more detailed description and idea behind the summarized solution.

As indicated in the summary section, we can decompose the solution into multiple parts. One is classifying who will pay (payers) and will not pay (non-payers). Another is determining which shutoffs should be performed given who we believe will not pay. We also use an estimation model for expected future payments, which is used to forecast the value (cost) of the shutoff.

### **Classifying Payers and Non-Payers**

There are many approaches to classifying payers and non-payers. For this exercise we'll use a logistic regression model which fits well with binary decisions and also has the nice property of yielding a probability. A probability is appealing in this case because we can couple it with an estimated payment and obtain an expected cost (saving) for each shutoff.

### **Payment Estimation**

For payment estimation, in addition to the classification task, we need an approach to estimating the expected payment. Note that we may not need data about current the current resident to estimate a payment since power consumption can be based historical data across all residents. For this type of model, one option is to use a simple arithmetic average based on historical data to estimate a payment for each residence for each month. this is easy to implement and easy to interpret.

### **Identifying Shutoffs to be Performed**

An optimization approach seems very natural to determine which shutoffs to perform. This is because we have a limited resource we need to utilize, namely the the number of workers and their time. We also have a cost in the form of expected payments.

From the classification part of our approach, we have locations of likely non-payers. From the payment estimation part of our approach, we have an expected value for a pay amount. We can couple these to differentiate high value locations from low value locations.

We can use an optimization approach where a worker is assigned to cover a portion of the map where the goal is to maximize value. This is simliar to the optimization technique where we determine where to place warehouses to optimize the cost of delivery.

### **Data**

As pointed out in the week 8 lectures, laws at various levels of government may prohibit the use of certain data. Data likely to be prohibited relates to demographic attributes of the individuals at a residence, and also less obvious items that may be correlated.

Historical payment information and power consumption is likely available within the billing system of the power company. Worker availability and scheduling is also likely to be available within one of the company's information systems.

Current resident information may also be available within the company's billing system as such information may have been included when the resident was originally setup in the system.

Since we'll have the address of each resident, we'll be able to merge location information (e.g. longitude and latitude) with our expected payment information and also predicted non-payers.