Chris Messer
IYSE 6501
Power Company Case Study
11/12/22

Summary

Case Facts:

- An electric company (EC) needs to shut off the electricity for some non-paying customers.
- The company needs to determine what restrictions to apply regarding *which* customers to shut electricity off for. The company needs to consider the following:
  - Some non-paying customers will pay in the future
  - Some customers should be prioritized over others in the order they are shutoff.
- The EC has capacity restraints that apply to turning off electricity.
  - Travel time between each house to turn the electricity off.
  - The power is shut off using human labor, and a worker must be phycisically present to turn off the electricity. The company has a finite number of workers.

Component Analysis

To use analytical modeling to help the company, one should break down the components of the case into a mutually exclusive, collectively exhaustive set of problems to answer. Each component can then be analyzed and solved. These components can then be further disaggregated into sub-components, which we can use further modeling to I have identified the following components:

1. How should the company determine what customers they should shut off power to?
2. How should the company prioritize which customers to shut off, given there is a constraint on how many can be shut off per month?
3. How should the company optimize the physical process of shutting off power to these customers?

Solution

I have summarized the below analysis here for each of the above components.

1. **Given** customer information such as payment history, credit score, and residence type, **use** a flexible classification model combination of decision trees and logistic regression **to** identify which customers will not pay in the future.
2. **Given** the probability a customer won't pay, the cost of supporting the customer, and expected future revenue,
   use a linear optimization model
   to prioritize the order in which we would shut off the customers.
3. **Given** the customer's physical location, the cost of a worker, and the time to shut off electricity,

Use a clustering model to group the customers and an optimization model
To determine the route to drive and the number of workers needed

Analysis

Below I have broken each case component down, and will discuss the considerations, analytical approach, and data needs to solving each component of the case.

1. What customers should have their power turned off?

The case does not state whether the power company is a for profit or government run entity, but for purposes of this analysis, we will assume the company's goal is to maximize revenue and minimize costs.

The decision to turn off a customer's power is a binary decision; turn the power off or leave it on. Thus, we can simplify this question to align with the company's motive- if the present value of all future a customer's payments does not exceed the present value of all future costs to continue to service to the customer, then shut off the electricity. More simply put, if a customer will not pay in the future, turn off their services.

Analytical Approach

This part of the overall solution will likely be the most scrutinized part of the case as a whole- executive management would likely be too far removed from the day-to-day operations to be concerned with the order which customers have their power turned off or what route the trucks drive, but they will likely need a clear picture on how we determined which customers to shut electricity off to. For example, they would need to be confident that we are not turning off electricity to our biggest client because they were one day late.

Therefore, in this portion of the model, I would favor visibility and simplicity over accuracy. The last thing we want to do is put a black box-type model infront of management for a highly scrutinized component.

For this, I would use a combination of the following models:

- A decision tree to group customers into different buckets
- A logistic regression model within each terminal leaf to determine the likelihood a customer in that leaf node would one day pay

Then, I would use a cost matrix to determine the optimal boundary for which we should cut off a customer's electricity, i.e. at what percent likely a customer will not pay should we shout off their electricity? This would give us a clear number we could present to management. "For customers that meet this parameter, they are X% likely to never pay their bill. If we cut off electricity for customers who are Y% or more likely to not pay us, we will write off $z in receivables and lose $zz in revenue."

Data Needs

To build the analytic model portion (the decision tree and the logistic regression model), we would need to perform some validation on what the best indicators of non-payment would be. But also, we would need to determine what non-payment means. For example, if someone stopped paying yesterday, how can we know if they will pay in the future?

Defining non-payment

One approach to defining non-payment is a rollback technique. For this, we would look back at customer data from n years back, and if they have not paid as of today, consider them to not be paying ever.

Note: This approach does assume that the company does not have a policy in place to turn off electiricty after d number of days. For example, if they auto shut off anyone who has not paid in 90 days, then looking back two years would not be helpful—all customers who have not paid would in 90 days would not have paid after 2 years, because they stopped receiving service after 90 days! In this case, we would need a shorter roll back period.

The way this case is structured implies there is no company policy surrounding automatic shutoffs. Thus, we will proceed with a rollback. We will look back two years. We will look at all the attributes of customers from two years ago as the features for our model, and the response would be whether they paid within the next two subsequent years.

Feature Selection

For selecting which features we wanted to use, I would use lasso regression, as I would expect there to be many variables, and we are unsure which are relevant. Since lasso regression allows for 0 coefficients, we could eliminate unimportant metrics from our model. Some data points I would consider are:

- Income of the customer
- Zipcode of the customer
- Type of residence (Commercial/Resedential)
- Customer Occupation
- Time since last payment
- On-time payments over the life of the account
- Credit Score of customer
- Household size

There are somethings we would want to be aware of before diving in. There are certain protected attributes about a customer that we would legally not be allowed to use. For example, there are protections against using information like race, familial status, sexuality, disability status, etc. Even though we may have this information on file, and it may be a highly correlated factor to non-payment likelihood, we must be certain to comply with legislation that prevents the use of this information in our models.

After validating these factors with lasso regression, we could then build our decision tree model and then the subsequent logistic regression model on each terminal leaf.

Cost evaluation

Finally, once we have assigned a likelihood of non-payment, we would need to assign a cost for the cost matrix. For example, what is the cost to the business for shutting off a customer who may have paid? What's the cost of leaving on a customer's electricity who would not have paid? Once we have determined these costs, we could loop over the decision boundary to find the minum threshold of likelihood of non-payment, and that would be our final threshold. All customers who are XX% likely to not pay or higher will have their electricity shut off.

2. How should the company prioritize which customers to shut off first?

Assuming the company Is for profit, when evaluating which customers to shut electricity off to, the company's ultimate question is,

1. Which customers will bring in the least amount of revenue?
2. Which customers will cost the company the most money?

Here, we would need a method to rank the order of which customers we want to shut off electricity for. Because this part of the model would be less scrutinized, I would value accuracy over simplicity.

Analytical Approach

Because a highly accurate model is desired, I would use an optimization model here.

Objective Function

We would need to consider several things in our objective function.

- What is the potential revenue we may collect from this customer if we were wrong about them not paying? i.e. if it is a large customer with a large potential payout, and they were right on the decision boundary cusp of shutting them off, we would want them to be shut off last since the may pay.
- The total cost of supporting the customer until they are shut off. If that large customer is using a very high volume of energy, it may not be worth continuing to support them until we turn their electricity off.

The objective function we want to minimize would be

$\sum_{i=1}^{c}(R_i * n_i * p_i) - (C_i * n_i)$ = Ec

Where:
C = number of customers
i = customer

R = present value of revenue in perpetuity for each customer
p = likelihood to collect from that customer (determined in the first part of this case)
n = number of months until the customer is shut off
C = present value of the cost of service to that customer in perpetuity
Ep = expected cost for supporting that customer until they are shut off.

Variable Selection

The only variable we would be changing in this calculation would be n, the number of months until each customer is shut off.

Constraints

There would be a number of constraints we would need to consider. To make this function work.

- N must be positive, we can’t shut them off a month ago if they are currently active
- N may have an upper limit, if management has asked this be done in a certain time frame.
- If we can only shut off 30 customers a month, then we would need to link n to another variable, which would be the month the customer is shut off. This is similar to the food optimixzation problem. We used a variable in the constraint that was not in the objective function, whether the food was selected or not. We would need a constraint to say that there cannot be more than 30 customers in each “round” of shut offs

Using these features, we could get a list of all customers on the shut off list, and each customer would have an n value. An n value of one would mean the customer gets shut off in the first month, and an n value of 2 would mean the customer would be shut off the second month.

Data Needs

The data needs here are objective and easy to gather. We would need to know R, C, and p. R would be calculated from the net present value of their average use each month times the amount they pay for electricity.

C would be taken from their average use each month (we could get even more technical here and use something like a time series model to predict energy use over the year for each month they continue to use service) * the cost of electricity.

And p is calculated from the first part of this model, the decision tree and logistic regression.

3. How should we optimize the physical process of shutting off the power?

Lastly, we need to determine how we will actually go about shutting off the power to these customers. This can be broken down into two components:

1. How many workers do we need
2. Where should they be deployed to each month?
3. What route should they drive?

How many workers?

I won't walk though the full optimization model here, but I believe this would be an optimization problem to solve. We would need to know the cost of labor for each worker, the time it takes to shut off the electricity, and the number of customers to shut off. We would want to minimize the total cost of this process.

Where should they be deployed?

Here I would use a clustering algorithm, such as KNN. This would allow us to group customers together based on physical proximity to each other, and deploy n number of workers to each centroid of the clusters.

What route should they drive?

Fortunately, there are plenty of advanced optimization models here we could use that we would not even have to design ourselves. For example, shortest path algorithms could be used to determine the best route to drive to minimize time traveling.