Problem:

A power company needs to prioritize shutoffs for customers unwilling to pay bill.

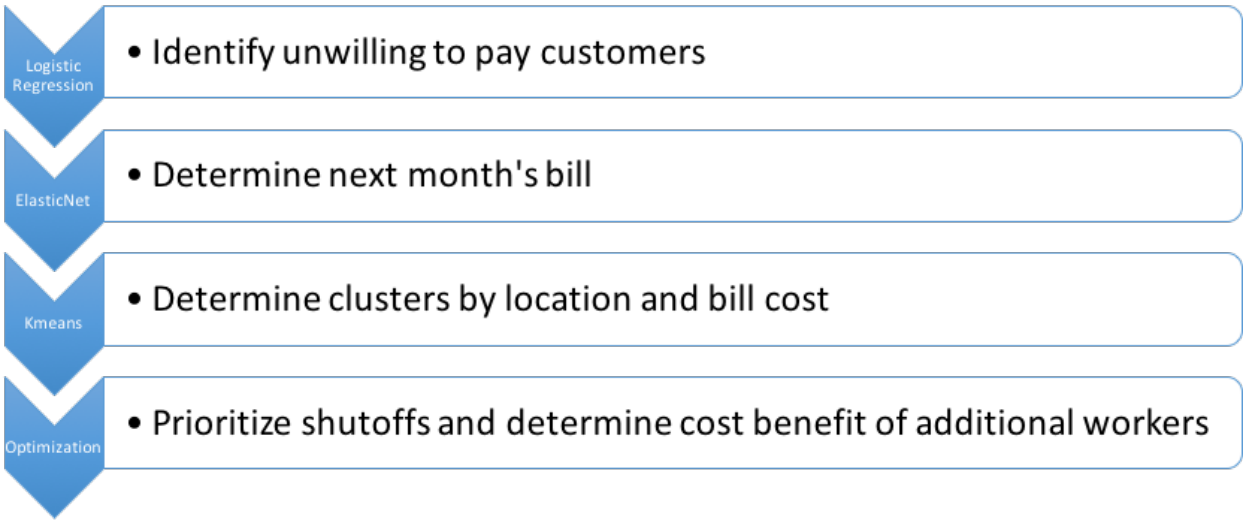
Assumptions:

- There are two groups of non-paying customers.
    1. Those who cannot pay
    2. Those who will not pay
- Not every shutoff is equal
- Shutoffs are constrained by the number of workers, distance between locations, time per shutoff

Given:

- Pay vs not pay
- Customer Credit Score
- Income
- Payment history
- Electricity consumption history
- Next month weather prediction/average
- Other seasonal affects (holidays)
- Customer location
- Number of workers per shift
- Average drive times
- Average shutoff time

Model Organization:



Logistic Regression

Given a set of non-paying customers, a customer's:

- credit score
- income
- payment history
- electricity consumption history

can be used to predict the probability that a customer is classified as unwilling to pay vs unable to pay with a logistic regression model.

ElasticNet

Given an identified set of unwilling to pay customers, a customer's:

- credit score
- income
- payment history
- electricity consumption history

As well as seasonal data such as:

- weather
- holidays or special events

An ElasticNet Regression model can be used to predict a customer's following month's electricity consumption and bill.

K-Means

Given an identified set of unwilling to pay customers, a customer's:

- predicted monthly usage
- location

Can be used to cluster customers both geographically and by cost savings by shutdown.

Optimization

Given a set of clusters, a cluster's:

- predicted monthly usage
- distance from work site
- average distance from cluster center

As well as historical worker data for:

- average drive times
- average shutoff times
- worker per shifts

An optimization model can be used to prioritize cluster shutoffs as well as determine the cost benefit to hiring any new workers to increase the shutoff capacity

Objective Function:

Maximize cost savings for shutting off power

Variables:

- Cluster shutoff (binary)

Constraints:

- Number of current workers
- Cost to train new workers
- Speed limit
- Shutoff time
- Hours per shift
- Overtime pay

Discussion

The outline above prioritizes the shutoff of unwilling to pay customers. It identifies unwilling to pay customers with a logistic regression model, predicts next month's bill with ElasticNet Regression and prioritizes shutoffs using an optimization model. I think the model is simple, and I could perform most of it with my current experience. Therefore, I think it is a solid, and not overly complicated method to prioritize shutoffs. It is not without some obvious shortcomings, however.

The majority of my problems are in the optimization process. I do not currently know how to create a vehicle routing model, which I think would be necessary to fully flesh out the optimization model and may obviate the k-means model. Although I think this would do an ok job, the k-means only presents geographic location, it does not tell me how easy it is to get from one location to another. For example, two locations may be grouped together because of their proximity, but are separated by a river with the closest bridge being several miles away.

I do not believe my optimization model is correctly stated above. I almost would need to optimization models. One to determine which clusters to shutoff and a second to determine the optimal number of workers. It may be possible to accomplish both by adding new workers as a variable.

Lastly, I do not have anything to determine what the probability threshold for my logistic regression model should be. I think there are more considerations than just cost for power. Such as neighborhood and societal costs what are these costs when we get it wrong, since these are probabilities. If a particular neighborhood is hit harder than others house values can go down. These are other interesting and potential factors to keep in mind.