POWER COMPANY CASE STUDY

HW 12 – KELLY “SCOTT” SIMS

Table of Contents

Payers vs. Non Payers

The first Analytical step, naturally, would be compiling a list of payers vs. non payers. A non payer is defined as someone who will never pay their bill even if they have the means to do so (no extreme economical hardship that prevents any payment). Given that in America, one cannot discriminate against: age, sex, race/ethnicity, or nationality, factors must be carefully chosen in this step of the Analysis. Even these factors might have strong co-linearity with other factors that allude to discrimination. E.g. zip codes, area codes, and street names could be clusters of similar groups of the aforementioned factors. Therefore:

Given:

- income
- credit score
- energy usage
- late payments to total payments ratio
- no payments to total payments ratio
- energy usage to income ratio
- residential or business encoding
- length of energy account

Use:

An ensemble "Majority Voting" algorithm to classify payers vs non payers. The voters will be:

- Logistic Regression
- Random Forest
- Naive Bayes
- SVM
- k-Means Clustering

Each voter makes a prediction and returns the classification probability (not the hard classification). All 5 probabilities are then fed into a:

- Neural Network with 2 output layers

Results:

The neural network returns a response 0 {payers} or 1 {Non payers}

Predicted Future Use

After compiling a list of predicted Non payers, the next step is rank those non payers in order of “Greatest Energy Users” to “Lowest Energy Users”. This is an exercise in finding those who in the future will run up the largest bill while maintaining a delinquent account. In order to accurately predict this, we must first break our offenders up into two categories:

- Customers with 2 years or more energy history
- Customers with less than 2 years energy history

Customers With 2 Years or More Energy History

Customers with 2 years or more energy history with the power company have presumably established a distinctive usage trend.

Given:

Historical energy usage of the Non payers

Use:

Exponential smoothing with forecasting

Results:

Future usage of the delinquent offenders ranked from most to least

Customers with less than 2 Years Energy History

It can be argued that less than 2 years history isn’t a reliable enough source. The first year in a new home/apt could have erratic usage due to moving, renovations, remodeling, or many other things people do their first couple of years in a home to get settled in. Because of this, forecasting might not be prudent. Since we’ve already separated our payers vs non payers avoiding illegal factors, we can now reintroduce those factors to at least estimate usage per area. We need factors such as zip code, street, and age because certain areas have bigger houses than others, meaning more usage. Younger people arguably have more tech and electronics that consume a lot of energy. Certain areas might be colder or hotter than others due to elevation or nearby bodies of water. Therefore:

Given:

- Age
- Upcoming Month
- Average Energy Usage For given customer’s zip code in the Spring (if distribution is normal, otherwise use the median)
- Average Energy Usage For given customer’s zip code in the Summer (if distribution is normal, otherwise use the median)

- Average Energy Usage For given customer’s zip code in the Fall (if distribution is normal, otherwise use the median)
- Average Energy Usage For given customer’s zip code in the Winter (if distribution is normal, otherwise use the median)
- Average Energy usage for given customer’s street in the Spring (if distribution is normal, otherwise use the median)
- Average Energy usage for given customer’s street in the Summer (if distribution is normal, otherwise use the median)
- Average Energy usage for given customer’s street in the Fall (if distribution is normal, otherwise use the median)
- Average Energy usage for given customer’s street in the Winter (if distribution is normal, otherwise use the median)

Use:

Linear Regression

Results:

Predicted future energy usage

We then combine the results from both groups and the aggregation results in the descending order of highest predicted energy users to the lowest.

Network Analysis

Using our lists of delinquent users we can perform a Social Network Analysis and cluster the delinquent customers by their address. Addresses nearest to each other, obviously will form a cluster. Each node in a cluster (delinquent customer) will be color coded by its predicted energy usage. Each cluster as a whole will be color coded by the areas that are predicted to consume the most energy.

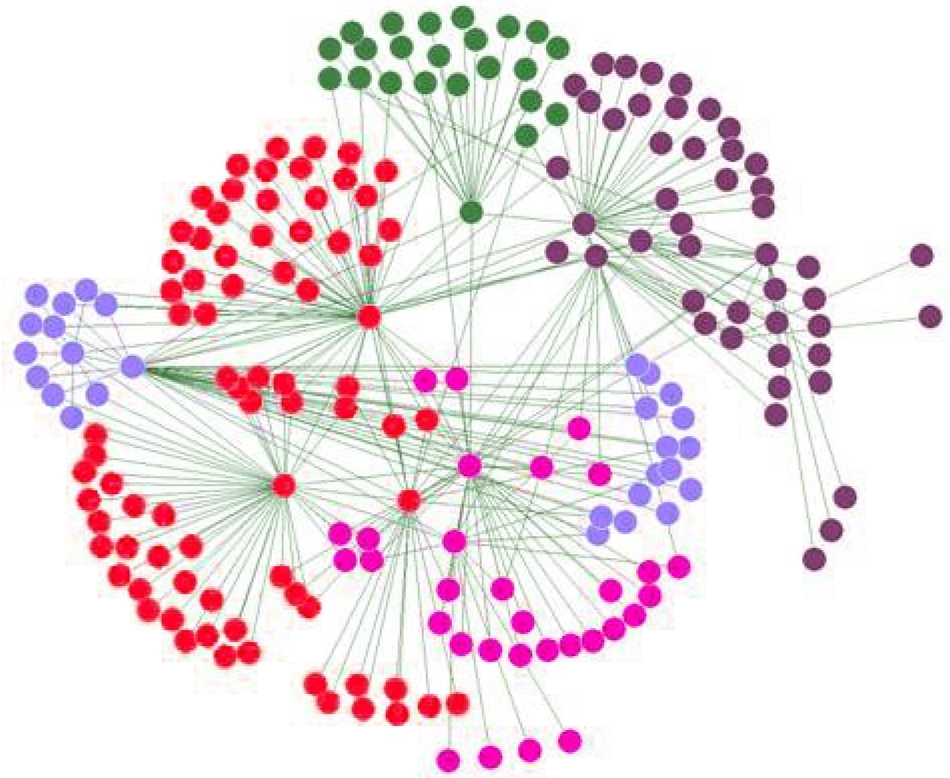

Illustration 1: Example of a Social Network Analysis Graph. Not exactly representative of the method being described

Therefore:

Given:

- Customer’s addresses
- Customer’s predicted energy usage

Use:

- Social Network Analysis

Results:

- Clusters throughout the city depicting areas of high to low predicted energy use

Optimization Model

Now that we have our customer’s grouped in distinct clusters, we know exactly how much energy is predicted to be consumed with respect to each clusters, the square mileage of each cluster, and the projected revenue loss per cluster. We can now map out our plan to being shutting off power. Treating each cluster independently, The first thing we need to figure out for the optimization model is the routes in each cluster to maximize shutoffs while minimizing drive times as well as employee times. Once we have this information for each cluster, we can begin the optimization.

Vehicle Routing Problem

Given:

- Addresses within each clusters
- Ability to vary the number of employees in each cluster to shutoff power

Use:

- Recursive-DBSCAN (vehicle routing problem algorithm)

Results:

- Optimal routes per cluster per amount of employees in each cluster with respective drive times

Final Optimization Model

Armed with all our data, we can begin the optimization model

Given:

- Cost of gas
- Average Miles per gallon of employees trucks
- Hourly pay of employees
- Route times within each cluster with respect to amount of employees assigned to each clusters
- Estimated time it takes to shutoff power
- The predicted amount of energy used per address in each clusters
- Maximum allowable hours worked by a single employee

Use:

- Optimization to:
  - Minimize operational costs (Hourly pay + [cost of gas per gallon x (total mileage / average mpg)])
  - Maximize number of shutoffs
    - Constraint: Hours of employees cannot be greater than maximum allowable hours

Results:

- Number of employees needed for each clusters
- Planned route

- Maximum number of shutoffs resulting in the greatest reduction of energy consumption

Alternative Solution

Instead of using an optimization model, we can use simulation. Knowing the optimum planned route for each cluster with respect to the number of employees in each cluster, we can run multiple simulations, varying the amount of employees in each cluster. Therefore:

Given:

- Optimum route in each cluster with respect to amount of employees in each cluster
- The average time it takes to shutoff power
- Time of rush hours in each clusters and average delays due to rush hour
- Maximum allowable work hours per employee

Use:

- Simulation and:
  - vary number of employees in each cluster
  - vary starting time of each employee so as to avoid having all employees working at the same time and/or during rush hour
  - End simulation once all employees have reached maximum worked hours for the day

Return:

- Results of each model:
  - Number of shutoffs
  - Total distance driven
  - Number of employees in each simulations

Use this data to calculate total mileage driven which is then used to calculate the cost of fuel, then umber of employees and the cost to pay them all, the number of shutoffs performed and how much energy was saved performing these shutoffs. From this data, the power company can just the optimal model that suits their business needs/plans