

HW 8

Possible Recommendations to the Power Company

Before we use any analytical models, it is vital to investigate the data that has been collected by the power company about its customers. Some of the critical factors to consider includes the following, credit score, income, what type of housing do they live in (apartment or house), past defaults, and past power bill payment history. This data should be compiled for the past few years (lets day 2) to see any trends/pattern in missing data. Based on these factors, the company could then use a classification approach to categorize its customers. Here customers are the people who is currently registered with the power company to pay the bill (that includes):

1. Customers who pay the bill regularly
2. Customers who occasionally forgets to pay the bill (but not always)
3. Customers who never pay the bill despite being registered to the power company

After categorizing its customers, the company should then see if there are any missing data and if there a pattern that follows the missing data. Since we are interested in the customers who have not paid their bills, we should focus on investigating the missing data information for that category. Sometimes, there could be long term (but yet temporary) factors such as unemployment, medical issues etc. that can hinder an individual from paying their bills. The power company should compile data on how long an individual has not paid their bills (is it a month or two months or more) and how often is this pattern repeated. For example, John has not paid his bill for the past two months, but pays back his bill on the third month. But he does not pay his bill back again the fourth and fifth month but pays it back the sixth month. As you can see there is a bit of a pattern here. If there are any missing data for the below factors, below are possible solutions that could be used to estimate a value for missing data:

1. Credit score (Numeric) : Can use mean, median, or mode
2. Income (Numeric) : Use regression and perturbation
3. Type of Housing (Categorical): Use categorical variables to indicate "missing" data
4. Past Defaults (Categorical): Use categorical variables to indicate "missing" data
5. Past Power Bill Payment History (Categorical): Use categorical variables to indicate "missing" data

We can use the above factors to not only classify customers but also come up with an estimate of how much power the customers will use next month. The power company can use Exponential smoothing or regression to come with an estimate for that. These models will need to be validated and cross validated on training, validation, and test data to come up with a good estimate. Based on the pattern seen in the missing data and also the expected cost of leaving the power on or off, we can compile a priority list. Once we have a priority list of locations

where the we shut off power, we can create a cluster map. The cluster map will help the worker shut off power for large groups of locations that are clustered with each other.

In order to determine the time it takes to travel to the location in the clusters, we will need to compile past data on travel time and speed (mpg) of the vehicles that is used to do this. We can use optimization model here since we are trying to identify the highest set of customers for power shut off by using limited resources. We can use clustering as well. We can then rerun our model using simulation to manipulate our variables.