

Question 18.1

Describe analytics models and data that could be used to make good recommendations to the power company.

Here are some questions to consider:

- 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.
- 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?

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

Have fun! Taking a real problem, and thinking through the modeling and data process to build a good solution framework, is my favorite part of analytics.

Answer :

Case study : Power Company

In this case study, there are a large chunk of customers of the power company who are not eligible for Power Shut offs as those who will eventually pay or those who are eligible for government funding due to low income category.

Phase 1: Categorize the Power company customer data into two groups – “Customers who are able to pay the bills” and “customers who are unable to pay the bill due to low income (eligible for government funding)”

Given	Customer Name, Age, Education Level, Occupation, Current Employment Status (Have a job or no job), Prior Payment History, Length in delay of payments, Home Ownership, Zip code, Income, Financial debts, Credit Score, number of household dependents
Use	SVM classification.
To	Classify the Customers “who will be able to pay the bill” and “who cannot pay the bill.”

Note : Customers who fall under the low-income group and who might be eligible for funds from the government based on the low income or no income will fall into the categories that cannot pay the bill. Level of Education, whether holding a job or not etc. are all indicators to correctly classify the customers into the above two groups.

Phase 2: Get the data of the customers who will be able to pay the bill from Phase 1 classification and classify “whether they will pay the bill in the future” or “ will never pay the bill”

Notes : Phase 2’s input fall under average / above average income groups.

Given	Customer Name, Age, Education Level, Occupation, Prior Payment History, Length in delay of payments, Home Ownership, Zip code, Income, Financial debts, Credit Score
Use	Logistic Regression
To	Classify based on the probability of the Customers into “who will eventually pay the bill and who will never pay”

Phase 3: From Phase 2 – We have the Customer list who will not pay the bill. Now we can go the Cost estimation of for the coming few months to create a priority list for Power Shutoffs.

Cost Estimation if the Power is left on:

Given	Customers past power usage data, weather information for seasonality (High summers and Winters will usually have heavy power consumption)
Use	Holtswinter method for exponential smoothing with seasonality and ARIMA
To	To forecast the customers upcoming months power usage. And create a priority list of High-power usage customers and are eligible for power shut off.

Phase 4: Decide whether the Customer’s power should be turned off?

Given	For Customers from the High-Power usage list of people who will never pay)– Number of technicians available and the cost of doing the shut off (travel time , labor cost, How many nearby places of shutoff), location address (zip code) , Traffic and Navigation Data
Use	K means (to cluster the locations)
To	Get the clusters of customers with similar chance of shutoff and locations.

Phase 6: Minimize the Shutoff costs.

Data : From Phase 5 : clusters divided based on location and who are eligible for possible shut off

Given	Number of technicians available and the cost of doing the shut off (travel time , labor cost, How many nearby places of shutoff), location (zip code) of each cluster center , Traffic Data and Navigation data
Use	Optimization to minimize the shut off cost. Using the Constraints like more than a threshold of customers must be located in the same area to perform shutdown.

To	Get the clusters of customers with similar chance of shutoff and locations.

Conclusion :

First the customer data is classified using the SVM model into “Customer can pay” and “ Customers cannot pay because of low income” .

Next using the data of Customers who can make the payments,
Using Logistic regression using the probabilities , the customer that can pay are further classified into those who will eventually pay and those who can pay but will never pay- eligible for shutoffs.

Then using Exponential Smoothing (Holtswinter method) and ARIMA on the “Shutoff eligible list of customers” , the cost of high-power usage (negatively impacting the Power company’s revenue) for upcoming months using seasonality’s is estimated and a priority list is created for Shutoffs.

Then using K means location clusters are formed and using Optimization to minimize the shutdown cost using the constraints like more than a threshold of customers must be located in the same area to perform shutdown.