Marcus McKenzie
Analytics Modeling
Power Case Study
November 11, 2020

Given: {Customer Info (payment history, family size, age, income, etc.), Consumption, External factors(weather, cost of energy, seasonality)}

Use: {classification model - random forest, SVM, boosting}

To: identify customers who are likely to never pay

Given: {Customer location, probability to never pay}

Use: {K-nearest neighbor clustering}

To: {determine areas where groups of customers have increased probability to never pay}

Results:

The result would provide a recommendation to the power company each month about where to send the workers and to which locations that would minimize the least cost for the power company. In addition a message could be sent out to each of the customers that are expected to not pay the bill. This may help to reduce and determine the amount of customers that will not pay as opposed to having forgotten to pay. This additional information to help us decide which clusters, and how many workers are sent when turning off the power.

Data:

Above, we described data points the energy company would most likely have access to, along with exogenous factors that might impact the probability of a missed payment. Customer payment history and factors that provide information about a customer may help to accurately identify customers at risk. For example, younger customers that may have inconsistent payment history may be most likely to never pay. However, families that have higher incomes may be more likely to eventually pay off their debt.

In order to effectively classify customers we will need many examples of customers who default completely and customers who default for some period of time then pay their remaining balance. For this, a certain threshold or amount of time will have to be set to determine that a customer will actually never pay. In theory, a customer labelled as unwilling to pay, could pay the next month.

Model:

A classification tree model will allow us to draw quick insights by viewing the decision tree splits and classification scores for each classification. This in turn, would allow for the use of a Random Forest Model. This would then allow us to fit the data, between types of non-paying customers. Both of these tree based models can determine the probabilities that a customer will not pay. Using probabilities as opposed to other classification methods would allow us to determine the company's desires based on their available resources. By removing the classification threshold, we can select shutoffs for customers who are not likely to pay or lower the threshold depending on the resources set aside.

By determining the probability we can make informed decisions about who to send messages to, to attempt to influence customers to pay. We could email customers across different ranges of non-payment probability and identify which are the most likely to be responsive. Hopefully, this will spur a number of customers to pay the balance, so as to not have their power shut off. This should limit the number of shut-offs and could also be used as a factor for future forecasting.

Optimizing

Reducing the solution to a problem about how to optimize shut downs for a specific cluster, would provide greater efficiency in the problem. Given a cluster we could use a shortest path algorithm(Djiskra's) to find the most efficient manner for the shut-down in a cluster. The number of workers, the best path and the order of the shut-downs could be provided to most efficiently determine the probability of likely defaults and costs of customer clusters.