HW #12

Question 18.1

The overall objective of this problem is to identify which customers have the capability to pay their bill but will not do so and how to best strategize the termination of power for those clients.

The first step is to identify the data that is required. As explained in the lectures, I do not want to focus the data collection on items that could be directly related to race, and any other factors restricted by regulators. The following are the relevant data items that I would be interested in:

- Age
- Sex
- Marital Status
- Credit Score
- Annual Income
- Number of people in the household
- Average monthly power bill
- Average amount of energy consumed/month
- Account balance
- No. of bills missed/not paid in the last 12 months
- No. of consecutive bills not paid in the last 12 months (this would help me to see if the customer has a pattern of paying once and not paying for many months and then paying a small amount later)
- No. of days since the last payment
- Own or rent
- House price/value
- How long has the customer lived in the current address/home
- No. of times the customer has moved in the last three years
- No. of times the customer has been arrested/served citations in the last 10 years (this would be hard to get but with the social security number should not be too difficult. This will help identify if the customer exhibits a similar pattern in other aspects)
- How long has the customer been employed in the current job
- No. of follow ups/remainders to the customer in the last 3 months
- Has the customer requested additional time to pay their bill?
- Has the customer filed any disputes or disagreements regarding the bill (incorrect usage/rates) in the last twelve months.
- Has the customer signed up for e-bill or do they get paper bills?

As shown above my focus on the data collection is mainly about the type of customer and their relationship with the power company.

Once I have all the data, the next step is to remove any outliers and use imputation techniques appropriately to infer missing data.

I will then use a classification model such as SVM or KNN to determine if a customer will pay their bill or not. However, the challenge is to separate customers who can pay their bill but will not do so from

customers who are not able to pay their bill due to other reasons. Therefore, I can take the results of the classifier above and filter the data to obtain the customers who will not pay their bill. I can then run this through a clustering algorithm such as K-means (k value of 3) to cluster the customers into three distinct groups - who can pay their bill but will not do so, customers who are not able to pay their bill due to financial reasons and customers who will not pay their bill due to other reasons such as meter reading disagreement or other disputes. Once I have clustered the non-paying customers into three different groups, I will focus on my target segment – who can pay their bill but will not do so.

To me these are the customers who would be worthy candidates for termination, provided there is an economic value of doing so.

As a next step, I will forecast the next six month's energy usage and the approximate bill amount for the customers in my target cluster as mentioned above. I will use exponential forecasting techniques that include seasonality and trends to predict the future energy usage and based on the usage I will determine their potential bill amount. This is to determine if it is economical for the power company to shut the power off for those clients.

Next, for each customer that I have determined as someone who will not pay their bill (my target cluster), I will find out the cost to terminate the power and the revenue that I think the customer will owe the company (above). The costs to the power company to terminate a connection could involve both fixed and variable costs. Once I identify the customers whose service is worth terminating depending upon the costs and anticipated revenue, I will create a discrete event stochastic simulation model to identify the best schedule for doing so. In addition, simulation can also help me identify how many resources/workers I need to accomplish this. Simulation also will help me decide on the priority of carrying out the terminations which is basically my schedule.

Specifically, I will create a simulation model to capture the power station workers travelling to different locations to terminate the power. I will run the model for different routes/schedules where the entities would travel to different locations which would indeed represent the locations of the customers whose power is supposed to be terminated. I will create resources and use them at different locations for a certain amount of time. These resources would represent the power company workers. It must be mentioned that I will use probability distribution for processing time etc., the processing times could be a triangular distribution. I can even model breakdowns or delays in travelling to the different locations and use a Weibull distribution.

I will create attributes to track the costs and assign to the entities. My overall objective is to devise a schedule that the workers can follow to terminate the power of all the customers at a minimum cost. In addition, I will also identify the number of workers I need to accomplish this in a limited time frame. So, I will track the total cycle time for each termination and strive to keep the termination time also to a minimum. Finally, I will leverage the optimization capabilities in Arena to model different scenarios and select the best one – i.e. I am looking for something like travel to location A, B, D, E, F and C to terminate power, hire three more resources, this would cost x dollars on average and take y minutes on average.