Homework 12: ISYE 6501 - Introduction to Analytics Modeling

Question 18.1

Prompt

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.

Approach

The case involves a power company dealing with customers who do not pay their bills. The company is not concerned with those who cannot afford to pay, but instead, the focus is on customers who can pay but choose not to. The power company wants to turn off the power for these non-paying customers but faces challenges in identifying those who will never pay versus those who might have just forgotten. To address the power company's issue of identifying which shutoffs should be done each month, we can use various analytics models and data.

With the use of company and customer data, predictive modeling and optimization modeling will be performed. A binary classification model is selected for predictive modeling to predict the likelihood of a customer eventually paying their bill if not shut off. This model will help identify which customers are likely to pay and which are not. To optimize the problem, the logistic regression model predictive results are utilized with a PuLP Linear Programming (LP) model. The optimization problem is formulated to maximize the number of shutoffs within the capacity constraints, considering worker time and travel. Constraints are also defined for prioritization of shutoffs based on the predictive model's output.

In result, prioritization is implemented to ensure a cost-effective and efficient model. Prioritization is implemented by multiples considerations; (1) Using the logistic regression model's probabilities to rank customers by the likelihood of non-payment, (2) Prioritize shutoffs for customers with the highest likelihood of non-payment, and (3) Incorporate operational constraints to schedule shutoffs efficiently.

Predictive Modeling

Predictive modeling can be utilized to predict future outcomes based on historical data. For the power company's case, the goal is to predict which customers are likely to pay their bills eventually and which are not, to make informed decisions about power shutoffs.

Problem Definition

The objective of the power company's problem is predicting the likelihood that a customer will eventually pay their bill if not shut down. The target variable is binary, determining whether a customer eventually pays (1) or does not pay (0). A prediction model should be developed for this binary classification problem.

Data Preparation

Predictive models require data preparation prior to executing a chosen algorithm. Models must consider feature selection, where relevant features need to be identified that will assist in predicting the target variable. The target (dependent) variable will be binary, indicating whether the customer will eventually pay (1) or not pay (0). The predictor (independent) variables will be historical data for the customer and power company. Predictor variables should include account history, outstanding balance, demographics, and historical data. Account history can include average balance, number of late payments, average payment delay, and payment consistency. Demographic information could include the property location and income for that income area. Historical shutoff data could include number of shutoffs and length of time power was terminated.

Model

Model selection is important so that the appropriate algorithms are implemented to identify solutions to the problem. Selection of a model must consider the power company's case and determine a problem that can utilize analytics to provide a specific solution. Considering whether a customer will pay their bill even if the power is not shut off, a predictive model that is appropriate for a binary classification problem must be selected.

A logistic regression model is selected for binary classification, which is a popular choice for simplicity and effectiveness. Logistic regression is beneficial to implement because it will provide probabilistic outputs. Probability results will allow for decision making on whether the customer will pay or not. Logistic regression can serve as the baseline model to compare with more complex models if needed in the future.

The logistic regression model will be trained on historical data to evaluate the relationship between predictors and the target variable. The model will also output coefficients for each feature/predictor, which are useful to indicate the impact of each regarding the likelihood of payment. Dependent on the input predictors, the model will estimate the probability that a customer will pay their bill. The model can be tuned by determining important predictors and only using that data in the logistic regression model.

Model Output and Results

The output of the logistic regression model will contain predicted probabilities to prioritize customers for shutoff. Customers with lower probabilities of payment are prioritized for shutoff, and the classification output of whether the customer will pay or not helps make binary decisions on whether to shutoff a customer's power. Probabilities and classifications can be integrated into the optimization model to schedule shutoffs efficiently.

Optimization Model

An optimization model is beneficial to implement since it aims to determine the most efficient and effective way to schedule power shutoffs, given the constraints and objectives of the power company. Optimization occurs by minimizing or maximizing the objective function subject to constraints.

Problem Definition

The objective of the optimization model is to maximize the number of effective shutoffs within the available capacity. The primary objective is in an effort to minimize operational costs and determine valuable shutoff decisions.

Data Preparation

To form a useful optimization model, the outputs from the logistic regression model will be implemented. The predictive model results that can be utilized for optimization models are probability scores for likelihood of the customer paying and risk scores based on payment behavior and factors for each customer.

As for historical data, customer data is described as used in the logistic regression model. Customer data includes account history such as payments and outstanding balance, demographic information and historical shutoff data. Operational data is important to consider for worker schedules, availability, travel times, and distance between customer locations. There also must be constraints for considering capacity (max shutoffs per day/month), and costs for performing shutoffs (labor/transportation).

Model

An optimization model can be formed using the PuLP package for a Linear Programming (LP) model. LP optimization models are beneficial because the methodology is utilized to determine the best outcome, such as maximum profit or lowest cost. Considering we are determining whether to shutoff or not, then Binary Integer Linear Programming (BILP) is considered. BILP is a special case of Mixed-Integer Linear Programming (MILP) where the integer variables are restricted to be binary (0 or 1).

To create an optimal model in PuLP that schedules power shutoffs, the constraints need to be defined. Constraints include:

- Capacity Constraints: Limit the number of shutoffs that can be performed within a given period
- Travel Constraints: Minimize the total travel time and distance between customer locations
- Operational Constraints: Ensure compliance with legal and operational requirements
- Customer Constraints: Prioritize customers with a low likelihood of payment

Model Output and Results

The optimization model will provide useful information in the output on shutoff schedule. The model will output a prioritized list of customers scheduled for shutoff. The schedule can be used to plan field worker routes and schedules, ensuring efficient use of time for labor. This will assist in an optimal process for reducing costs with use of prioritization.

Prioritization

Prioritization is the process of determining the order in which power shutoffs should be performed. Determining the order is based on various factors such as the likelihood of payment, and operational efficiency.

Prioritization uses both the predictive model and optimization model. The data from the optimization model is used, which includes customer data and operational data. The prioritization criteria will all be assigned weights for importance, to develop a weighted scoring system. The goal is to calculate a composite score for each customer based on criteria. The composite score will rank customers for shutoff. Criteria for analagous composite scores should be considered as well for shutoff (e.g., In the event of a tie, customer with highest outstanding balance is shutoff). This will allow for the power company to develop a robust prioritization process that ensures the most critical shutoffs are performed first, optimizing financial outcomes and operational efficiency while minimizing customer impact.

The result is an output of optimized routes and schedules for field workers to minimize travel time and costs. Outputs that also can be beneficial include the total number of shutoffs performed within the given period and total travel time and distance covered by field workers.