PowerCo Hypothesis Approach

Elham Ali

Email

Hi Maria,

I am following up on our initial team meeting regarding how to best approach the hypothesis that customer churn in the SME segment is primarily driven by price sensitivity. Below is my proposed methodology:

Hypothesis

Our primary hypothesis is that SME customers are leaving PowerCo because of price sensitivity. This means that we are looking to predict the likelihood of churn based on price-related factors.

- The outcome variable will be `churn`.
- The predictor variable will be `price` or `price changes`.

Data Required

To test this hypothesis, I would need historical data from PowerCo, specifically related to the SME customer segment. An ideal data frame would look like the following:

- **CustomerID**: Unique identifier for each customer.
- **Churn**: Yes/No (or 1/0).
- **Monthly_Bill**: Amount of the last month's bill.
- **Price_Change**: Percentage change in bill amount compared to the previous period.
- **Contract_Length**: Month-to-month, one year, two years.
- **Customer_Satisfaction**: Ratings or survey results.
- **Usage_Pattern**: Average usage over some time (e.g., monthly, quarterly).

- **Discounts_Received**: Past discounts offered.
- **Demographics**: Industry, company size, etc.

Key Steps

Below are the major steps I would do to test this hypothesis:

1. **Collect and clean data:** to make sure data is formatted correctly.
2. **Perform exploratory data analysis:** to understand the summary statistics, visualize the distribution of the data, correlations, and learn about patterns specific to SME customers who have churned. I plan to use R for this.
3. **Do feature engineering:** to derive features that may help in the predictive model like average percentage change in monthly bills, frequency of past discounts, etc.
4. **Build predictive model:** using machine learning algorithms to predict customer churn. Logistic regression would be a good starting point because of its interpretability and fit for binary outcome prediction. I could explore more complex models like Random Forest or Gradiant Boosting based on initial results.
5. **Evaluate model:** by using metrics like accuracy, precision, recall, and F1 score to assess the model's effectiveness. I would focus on the reliability of the model in predicting churn specifically due to price factors.
6. **Simulate discount strategy:** and use the model to simulate the effects of a 20% discount on customers most likely to churn. I would assess the cost-effectiveness of this strategy.
7. **Deploy:** once validated then we could use the predictive model on the 1st working day of every month to identify customers who should be offered a 20% discount.
8. **Monitor and update:** constantly by feeding new data into the model to keep it current and adjust for trends.

I recommend we start by first requesting the data suggested above from PowerCo this week so we can begin exploratory data analysis to identify the factors that contribute to SME customer churn.

Please let me know if you have any questions, comments or feedback. Thank you.

Warmly,

Elham

Additional Information

The below information will not be added to the email. However, it provides context behind my approach.

Null Hypothesis H_0

There is no relationship between price-related factor (like 'price' or 'price changes') and the likelihood of SME customers churning from PowerCo.

Alternative Hypothesis H_A

This is a relationship between price-related factor (like 'price' or 'price changes') and the likelihood of SME customers churning from PowerCo.

A logistic regression model equation for this hypothesis

- `exp` is the exponential function.
- The `coefficients` indicate how much the feature variables (like Monthly_Bill and Price_Change) affect the probability of churn.
- A positive coefficient means an increase in that feature increases the likelihood of churn. A negative coefficient means the opposite.

$$\text{Probability(Churn)} = \frac{1}{1 + \exp(-(\text{Intercept} + \text{Coefficient1} \times \text{Monthly\_Bill} + \text{Coefficient2} \times \text{Price\_Change} + ...)}$$