

Dear Team,

I wanted to share my thoughts on how we can approach and test the hypothesis that customer churn is driven by price sensitivity in order to gain valuable insights for PowerCo. To test the hypothesis, we can frame it as a binary classification problem. Given certain customer characteristics and product price data, we want to predict whether a customer is likely to churn.

Therefore, PowerCo must first provide us with historical data about customers, including their demographics, usage patterns, historical prices, and churn status.

We then perform feature engineering to create relevant features from the provided data that could influence churn. Further data preprocessing may be necessary to deal with things like missing data, outliers, or collinearity.

Next, we compare different ML models for binary classification, such as logistic regression, random forests, or gradient boosting. We train the models and evaluate them on the validation set using appropriate metrics such as accuracy, recall, precision, or f1 score. If the metrics are reasonably good, we can confirm the hypothesis. If the model performs almost randomly, the relevant features may not be included in our analysis. In this case, we can reject the hypothesis that price sensitivity affects churn and explore other features.

It is also important to be able to interpret the model's predictions to see if it is actually prices or other features that are driving churn. The final model is selected after considering the performance, complexity, and interpretability of the various models tested.

The trained model also allows us to backtest the proposed 20% discount strategy.

As the project progresses, we need to determine how the model will be deployed and made available to the client. As there may be data drift in the data used, the performance of the model will need to be continuously monitored and further action, such as retraining, will need to be taken if necessary.

Best regards,
Adrian