

REPORT ON PHARMACITCAL SALES CHURN CLASSIFICATION

We will try to predict whether a pharmacy customer makes a purchase in the month of October using data we have from the preceding months.

We will make use of the data from April 2019 to September 2019 as the training set and the October data as my test set. Also, we assign our target variable called "flag" a value of 1 if an October purchase is made and 0 if an October purchase is not made.

MY APPROACH

Organization of data: Each observation represents a single purchase at the pharmacy, so an extra time is spent organizing the data so that each observation is rolled up to the customer level

Selecting a model: I tried at least 9 different model on the dataset to visualize the predictive power as well as the level of accuracy.

POST ANALYSIS

We compared multiple models using a 5-fold cross validation. It is important to note that I did not use the test set (or holdout) with the actual October values. These will be used when evaluating the chosen model

DECISION

We will choose the model that contains the largest accuracy, which based on the LGBM Classifier. The accuracy reflects the amount of correct true and false predictions made and therefore is a good representation of the overall performance of the model. in this notebook we aim to predict whether a customer will make a purchase in the month of October, based on this we can explore other models to choose. If the pharma company's goal is to prevent lapse and thus identify those who will not purchase in October, then we might consider choosing a model based on the specificity which I did not calculate. On the other hand, if we were only interested in having the model maximize the number of predictions that predicted October correctly (ignoring which predictions the model accurately predicted not an October sale) we could focus on the Recall. Nonetheless, we will move forward with the LGBM or Light Gradient Boosting Model.