

TURNING APPREHENSION TO RETENTION

IL&FS Technologies turned customer apprehension to customer retention by using advanced analytics practises for a leading Indian Bank

Background and Scope

The bank who pioneers among Public Sector Banks in India in the use of Information Technology for Banking, has been noticing a quarterly decline in the CASA account numbers in the past few months. The bank wishes to carry out data exploration using Machine Learning to understand factor(s) that are affecting CASA Customer Churn, and also to develop a predictive model to assign a score to each of the high net worth customers, based on their propensity to churn. IL&FS Technologies approached the bank to carry out this exercise as a Pilot and illustrate the use of Machine Learning in solving such business problem.

The scope to IL&FS Technologies was limited to developing the Statistical Model that provides Probability to Churn of Accounts type Savings only; for the identified High Net Worth Individual (HNWI) customers. The team at IL&FS further specified the scope as per the geography / location, customer selection, QAB consideration and the transaction time period.

Sample Details

Majority of the data had three types of categories for the constitutional field - Individual, Joint and Others. The rate of churn was varying across all three categories Individual as a category was prominent and thus IL&FS conducted the exercise on this category only Furthermore, only standalone accounts with non-missing date of birth were considered. This was done so as to adhere to the definition of Churn (whereby a customer completely dissociates himself from there bank)

Solution

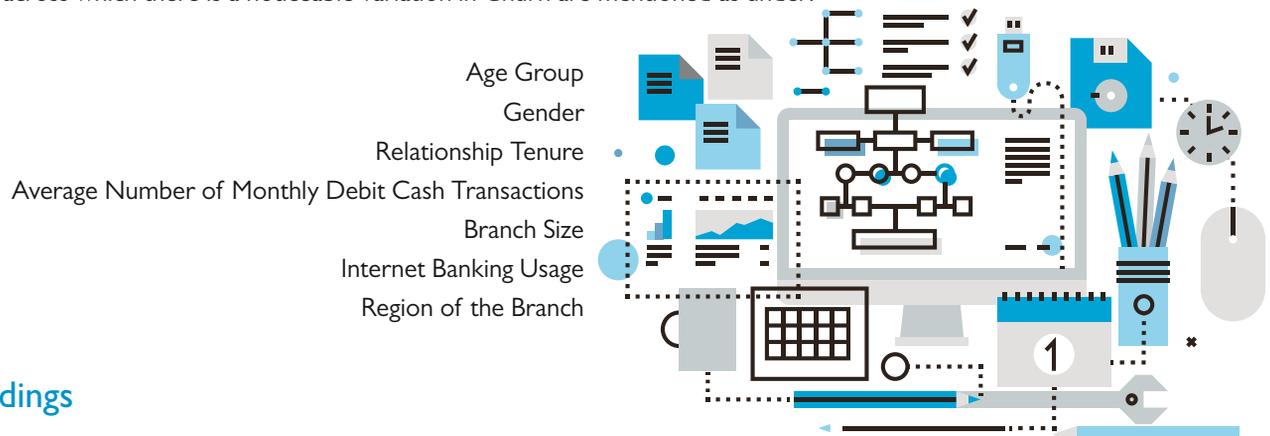
The entire dataset was split into training and testing datasets using an 80:20 split.

A logistic regression classifier was developed with 6-Fold cross validation. The topmost contributor to Churn was found to be the tenure of an account grouped into multiple buckets. Additionally, L2 normalisation was used so as to control the magnitude of the coefficients from taking on high values. Statistical / Machine Learning Techniques Explored: The problem at hand is a binary classification problem and the statistical techniques applicable are:

1. Tree based (Decision Trees, Random Forest, Boosted Trees)
2. Logistic Regression

Due to the highly imbalanced nature of the data and the prevalence of continuous features, decision trees yielded poor results for prediction. Hence, binary logistic regression was the choice for the model development process

Univariate data analysis was carried out to get an insight into the factors that contribute towards Customer Churn. Parameters considered for univariate analyses, consisted of variables related to customer as well as transaction. The parameters across which there is a noticeable variation in Churn are mentioned as under:



Final Findings

The primary outcome of the “Propensity to Churn” solution was a model that provided 'Churn Score' for the HNW customers. This helps the bank use this score to identify the list of likely customers who would churn. The bank can then easily contact them (through branch, email, customer care, phone etc.) to try to stop them from churning. By doing this, the bank reduced the churn and retained many high value customers. The study also provided the bank with insights on the factors that affect and cause a customer to churn. Similarly the study also provided the top factors that are noticed in stable customers.

The combination of these (factors that cause churn and factors that lead to long term relation) provided the bank with valuable insights on their customer behaviour and allowed them to streamline their operations.

The advanced analytics practise of ITL was able to segregate and identify the tenure, age group, QAB, size of the branch, demographics and behaviour of using the debit and credit cards - both for the customers who would churn and would most probably retain.