Project Details

Title: Beyond Churn Analysis: Retail Customer Churn Prediction in a Leading American Wholesaler with the Retailer's Personalized Insights - an Observational Study

Project members- MC. Hung, S. Dutta, V. Gaekwad, W.Sriapha, Y.Wang

Motivation

Reducing the rate of customer churn is a key business goal of every retailer. The objective of the study is to statistically analyze accessible historical data in order to discover patterns of demographic characteristics and shopping behaviors that promote customer churn. The accuracy of the predictive model is critical for preventive actions.

Dataset

The data was extracted from a secure source, with deidentified customer information such as account number, shopping information, transaction details, etc. The dataset had 120,450 rows. (Contact any of the above study members for further details)

Tech/ framework used

R Studio (Desktop version), R programming language
Tableau Desktop, Tableau Prep-Builder, Basic SQL queries

Method

When a straightforward implementation of a single method did not lead to a good performance, multiple methods were adopted to achieve the best-attempt results. The proposed models came in the form of individual models and models that utilized information carried forward from a different model. Missing data, multicollinearity, and data transformation were addressed prior to constructions of any models.

Files included

There are two 'Rmd' files, one with the codes for data visualizations, association rules, and the other with machine learning classification models. One tableau dashboard file also included.

- DM_project_EDA(Asso).Rmd
- DM_Porject_Models.Rmd
- Beyond Churn Analysis.twbx

Last updated: 24 April 2020

How to use?

1. To run the models, the dataset is to be uploaded in the IDE (R Studio)

2. 'DM_project_EDA(Asso).Rmd' should be executed first in the R Studio

3. Following packages are required to be installed to run the study successfully.
    DM_Porject_Models.Rmd should be run for model construction.

    library(readr)
    library(readxl)
    library(forecast)
    library(tidyverse)
    library(caret)
    library(e1071)
    library(data.table)
    library(randomForest)
    library(leaps)
    library (MASS)
    library(corrplot)
    library(ggplot2)
    library(cowplot)
    library(gridExtra)
    library(formattable)
    library(outliers)
    library(rpivotTable)
    library(InformationValue)
    library(ROCR)
    library(rpart)
    library(rpart.plot)
    library(FNN)
    library(arules)
    library(arulesViz)

4. Run the Beyond Churn Analysis.twbx file in Tableau Desktop

Contribute

The project studied human behaviors known to be challenging. The current data had some limitations due to factors such as a lack of date/time. The classification process can be improved by using advanced technique. Also, domain knowledge from experts in the industry may provide a better determination of variables used in the study.