Customer Analytics – Final Project

Due: 09.06. 2020 23:59 PM

Consider the data provided in "Final Project Data.xlsx" for 10,000 customers, who represent the cohort acquired in January 2019 (2019-1). Assume that there is no contract between the firm and the customers.

You are provided two sets of data:

- training data, covering periods between 2019-1:2019-9
- testing data, covering periods: between 2019-10:2020-2

Customers' transactions for a given month is represented with 1 (regardless of the number of transactions, as long as there is at least one transaction during the given month); and 0 otherwise (if there is no transaction during the month). Note that all data for 2019-1 is 1, as data set represents the customers, who made their first ever transactions in 2019-1.

Using training data, try out as many methods as you prefer (logistic regression, markov chain, time series etc) in order to predict the transactions during the testing interval (2019-10:2020-2).

Make sure that you include the probability based models (in addition to your models of choice) given below:

Probability Based Models:

a) assuming each year, every customer purchases with probability p

b) assuming each year customer i purchases with probability p_i

c) assuming each year customer purchases with probability p_i and churns with probability θ

d) assuming each year customer purchases with probability p_i and churns with probability θ_i

You are expected submit a Jupyter notebook (or a report in .pdf format). Please make sure that your report includes the following:

- Clearly stated assumptions and list of the models and formulas uses in the analysis.
- Presentation of results by using figures, summary tables etc. along with related discussions to increase the readability/understandability.
    - Make sure that you include analysis (graphs, projections, tables, accuracy results etc.) covering, at least, the following:
        - Cohort survival rates over months
        - Cohort retention rates over months
        - Prediction of total number of transactions per month
        - Prediction of cumulative transactions (over months)