MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT AND EBGINEERING

**Business Analytics Center of Excellence (BACoE): http://www.bacoe.in**

Faculty Evaluation for Data Mining and Machine learning

2016

Date of assignment: 16th April 2016



A super market is offering a new line of organic products. The supermarket’s management wants to determine which customers are likely to purchase these products.

1. The supermarket has a customer loyalty program. As an initial buyer incentive plan, the supermarket provided coupons for the organic products to all of the loyalty program participants and collected data that includes whether these customers purchased any of the organic products.
2. The ORGANIC dataset contains 13 variables and over 22000 observations.

The variables in the data set are shown below with the appropriate roles and levels:

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Model Role | Measurement Level | Description |
| ID | ID | Nominal | Customer loyalty identification number |
| DemAffl | Input | Interval | Affluent grade on a scale from 1 to 30 |
| DemAge | Input | Interval | Age, in years |
| DemCluster | Rejected | Nominal | Type of residential neighborhood |
| DemClusterGroup | Input | Nominal | Neighborhood group |
| DemGender | Input | Nominal | M = Male, F = Female, U = Unknown |
| DemRegion | Input | Nominal | Geographical region |
| DemTVReg | Input | Nominal | Television region |
| PromClass | Input | Nominal | Loyalty status: tin, silver, gold, or platinum |
| PromSpend | Input | Interval | Total Amount spent |
| PromTime | Input | Interval | Time as Loyalty card member |
| TargetBuy | Target | Binary | Organics purchased? 1 = Yes, 0 = No |
| TargetAmt | Target | Interval | Number of organic products purchased |

Though there are two target variables exists in this data set, let us consider only one target variable called TargetBuy (Binary)

Examine the data and answer following questions:

1. What is the proportion of individuals who purchased organic products?
2. The variable DemClusterGroup contains collapsed levels of the variable DemCluster. Presume that, based on previous experience; you believe that DemClusterGroup is sufficient for this type of modeling effort. Hence you may want to **not** consider DemClustor in your model building process.
3. As noted above, only TargetBuy will be used for this analysis and should have a role of Target. Can a TargetAmt be an input for a model used to predict TargetBuy? Why and Why not?
4. Devide your data to training and validation data sets as appropriate.
5. Build at least 5-6 models what you know and explain the case with a predictive model.

You have been given **two weeks time** to present your finding in front of a panel and demonstrate your on data handling, statistics, Analytics including model building machine learning and data mining. Explain about your model construction and model assessment process choosing the best model. Explain overall summary of this assignment.