**The level of customer loyalty**

ln this program, data is from Professor Jagdip Singh at Weatherhead School of Management from Case Western Reserve University. Programming by Jiaqi Li, Zeyao Yang and Ziming Cui.

***1. Usage Instructions***

If you wish to get a prediction result, change the dictionary paths for the data files containing the training and test data. Next, modify the output dictionary to point to the desired place. The software will output the predicted outcome.

***2. Background***

The level of customer loyalty will greatly affect the profitability of Bancare, so this program intends to collect customer data on LOYALTY and build a predictive model using a range of predictors as described in the “data-description” file.

***3. Methodology***

*To develop a predictive model*

1. Has little test error 2. A few best-chosen predictors were chosen using subset, ridge, or lasso regression. 3. Is trustworthy and unaffected by breaches 4. Contains interactions and terms from polynomials. Firstly, we completed data preprocessing and determining predictor correlation, and then the training set and test set are divided according to the ratio of 1:1. Ridge solution are not equivariant under scaling of the inputs, so we standardize the inputs before applying ridge. Secondly, Lasso can lead to zero coefficient which means some variables can be completely neglected. Divide the data at random into a training set (used to develop a predictive model) and a test set (used to evaluate the model). And then, calculate penalized logistic regression with Lasso Regression to automatically choose the best model with the most useful predictor variable. Finally, the graphic displays how each characteristic coefficient changes as lambda changes.

***4. Other Concerns***

Our model still exists multicollinearity between independent variables, which results in less reliable statistical inferences, therefore limit the research conclusions we can draw.

***5. Conclusion***

*Subset regression model*

Y = 0.2093 + 0.1186 \* Satis + 0.3026 \* VALUE + 0.1735 \* Reputa + 0.1183 \* CTRUST - 0.2337 \* educ4

*Ridge model*

Y = 0.0501 + 0.0739 \* Satis + 0.1992 \* VALUE + 0.20213 \* ATRUST + 0.0752 \* CTRUST + 0.1761 \* Reputa

*Lasso model*

Y = 0.2398 + 0.1210 \* Satis + 0.2875 \* VALUE + 0.1293 \* CTRUST + 0.1672 \* Reputa

The RMSE of Subset regression, Ridge, and Lasso model are 0.9019443, 0.9472489 and 0.8927231. As we can see, the RMSE of Lasso model is lowest, so this model’s accuracy is highest and we choose it as our best model.