COMMERCIAL BANKING, CORP

REQUEST FOR PROPOSAL
RFP #: IP – F1.H2

TITLE: BANKING INSURANCE PRODUCT – PHASE 2
CLOSING DATE AND TIME: SEPTEMBER 18. 2019 @ 5:00 PM

Banking Insurance Product – Phase 2: IP – F1.H2

Purpose

By responding to this Request for Proposal (RFP), the Proposer agrees that s/he has read and understood all documents within this RFP package.

Submission Details

Responders to this RFP should supply:

- A business report **up to 4 pages** (not including cover page, table of contents, or any needed appendix), including any supporting plots and tables.
- The commented code used to produce the results.

The report should address **all points described in the "Objective" section** below.

The report should be returned in the following way:

- Electronic (mailto: Aric_LaBarr@ncsu.edu; Subject Line: Banking Insurance Product – Phase 2)

Background

The Commercial Banking Corporation (hereafter the "Bank"), acting by and through its department of *Customer Services and New Products* is seeking proposals for banking services. The Bank ultimately wants to predict which customers will buy a variable rate annuity product.

A variable annuity is a contract between you and an insurance company / bank, under which the insurer agrees to make periodic payments to you, beginning either immediately or at some future date. You purchase a variable annuity contract by making either a single purchase payment or a series of purchase payments.

A variable annuity offers a range of investment options. The value of your investment as a variable annuity owner will vary depending on the performance of the investment options you choose. The investment options for a variable annuity are typically mutual funds that invest in stocks, bonds, money market instruments, or some combination of the three. If you are interested in more information, see: http://www.sec.gov/investor/pubs/varannty.htm

The project will be broken down into 3 phases:

- Phase 1 – Variable Understanding and Assumptions
- Phase 2 – Variable Selection and Modeling Building
- Phase 3 – Model Assessment and Prediction

Objective

The scope of services in this phase includes the following:

- For this phase use **only** the binned training data set.

- Based on your first report, the Bank has strategically binned each of the continuous variables in the data set to help facilitate any further analysis.
    - For any variable with missing values, change the data to include a missing category instead of a missing value for the categorical variable.
        - (HINT: Now all variables should be categorized (treated as categorical variables so no more continuous variable assumptions) and without missing values. Banks do this for more advanced modeling purposes that we will talk about in the spring.)
    - Check each variable for separation concerns. Document in the report and adjust any variables with complete or quasi-separation concerns.
- Build a **main effects only** binary logistic regression model to predict the purchase of the insurance product.
    - Use backward selection to do the variable selection – the Bank currently uses $\alpha = 0.002$ and p-values to perform backward, but is open to another technique and/or significance level if documented in your report.
    - Report the final variables from this model ranked by p-value.
        - (HINT: Even if you choose to not use p-values to select your variables, you should still rank all final variables by their p-value in this report.)
- Interpret one variable's odds ratio from your final model as an example.
    - Report on any interesting findings from your odds ratios from your model.
        - (HINT: This is open-ended and has no correct answer. However, you should get use to keeping an eye out for what you might deem important or interesting when exploring data to report in an executive summary.)
- Investigate possible interactions using forward selection including **only** the main effects from your previous final model.
    - Report the final interaction variables from this model ranked by p-value.
- Report your final logistic regression model's variables by significance.
    - (HINT: These steps are here to help you build your model, but **not** to tell you which order to write your report. Consider the most important information when done with these questions and write your report accordingly.)

Data Provided

The following two sets of data are provided for the proposal:

- The training data set **insurance_t_bin** contains 8,495 observations and 47 variables.
    - All of these customers have been offered the product in the data set under the variable **INS**, which takes a value of 1 if they bought and 0 if they did not buy.
    - There are 46 variables describing the customer's attributes **before** they were offered the new insurance product.
    - The Bank has strategically binned each of the continuous variables in the data set to help facilitate any further analysis.
        - (HINT: The original **insurance_t** and the new **insurance_t_bin** can be 1:1 row matched in case you wanted to know where the bins were split on.)
- The validation data set **insurance_v_bin** contains 2,124 observations and 47 variables.
- The table below describes the Roles and Description of the variables found in both data sets.
- (HINT: If you are using R, use the `haven` package and the `read_sas()` function to open the `.sas7bdat` files.

Name	*Model Role*	*Description*
ACCTAGE	Input	Age of oldest account
DDA	Input	Indicator for checking account
DDABAL	Input	Checking account balance
DEPAMT	Input	Total amount deposited
CASHBK	Input	Number of cash back requests
CHECKS	Input	Number of checks written
DIRDEP	Input	Indicator for direct deposit
NSF	Input	Number of insufficient fund issues
NSFAMT	Input	Amount of NSF
PHONE	Input	Number of telephone banking interactions
TELLER	Input	Number of teller visit interactions
SAV	Input	Indicator for savings account
SAVBAL	Input	Savings account balance
ATM	Input	Indicator for ATM interaction
ATMAMT	Input	Total ATM withdrawal amount
POS	Input	Number of point of sale interactions
POSAMT	Input	Total amount for point of sale interactions
CD	Input	Indicator for certificate of deposit account
CDBAL	Input	CD balance
IRA	Input	Indicator for retirement account
IRABAL	Input	IRA balance
LOC	Input	Indicator for line of credit
LOCBAL	Input	LOC balance
INV	Input	Indicator for investment account
INVBAL	Input	INV balance
ILS	Input	Indicator for installment loan
ILSBAL	Input	ILS balance
MM	Input	Indicator for money market account
MMBAL	Input	MM balance
MMCRED	Input	Number of money market credits
MTG	Input	Indicator for mortgage
MTGBAL	Input	MTG balance
CC	Input	Indicator for credit card
CCBAL	Input	CC balance
CCPURC	Input	Number of credit card purchases
SDB	Input	Indicator for safety deposit box
INCOME	Input	Income
HMOWN	Input	Indicator for home ownership
LORES	Input	Length of residence in years
HMVAL	Input	Value of home
AGE	Input	Age

CRSCORE	Input	Credit score
MOVED	Input	Recent address change
INAREA	Input	Indicator for local address
INS	Target	Indicator for purchase of insurance product
BRANCH	Input	Branch of bank
RES	Input	Area classification