COMMERCIAL BANKING, CORP

REQUEST FOR PROPOSAL
RFP #: IP – F3.H2

TITLE: BANKING INSURANCE PRODUCT – PHASE 2
CLOSING DATE AND TIME: NOVEMBER 18. 2022 @ 5:00 PM

Banking Insurance Product – Phase 2: IP – F3.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 (submit via Moodle)

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. Previously the bank sought consulting work on the same project, but also had a focus on understanding the factors involved. Here the focus is more on predictive power.

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 – MARS and GAMs
- Phase 2 – Tree-Based Models
- Phase 3 – Model Interpretation

Objective – Phase 2

The scope of services in this phase includes the following:

- For this phase use **only** the insurance_t data set.
- Previous analysis has identified potential predictor variables related to the purchase of the insurance product so no initial variable selection before model building is necessary.
- The data has missing values that need to be imputed.
    - Typically, the Bank has used median and mode imputation for continuous and categorical variables but are open to other techniques if they are justified in the report.
-
- The Bank is interested in the value of random forest models.
    - Build a random forest model.
        - (HINT: You CANNOT just copy and paste the code from class. In class we built a model to predict a continuous variable. Make sure your target variable is a factor for the random forest.)
    - Tune the model parameters and recommend a final random forest model.
        - You are welcome to consider variable selection as well for building your final model. Describe your process for arriving at your final model.
    - Report the variable importance for each of the variables in the model.
        - Pick one metric to rank things by – no need to report multiple metrics for each variable.
    - Report the area under the ROC curve as well as a plot of the ROC curve.
        - (HINT: Use the same approaches you used back in the logistic regression class.)
- The Bank is also interested in the value of an XGBoost model.
    - Build an XGBoost model.
        - (HINT: You CANNOT just copy and paste the code from class. In class we built a model to predict a continuous variable. You will need to look up the documentation for the ‘objective = "binary:logistic" ‘ option.)
        - Use the area under the ROC curve (AUC) as your evaluation metric instead of the default in XGBoost.
    - Tune the model parameters and recommend a final XGBoost model.
        - You are welcome to consider variable selection as well for building your final model. Describe your process for arriving at your final model.
    - Report the variable importance for each of the variables in the model.
    - Report the area under the ROC curve as well as a plot of the ROC curve.
        - (HINT: Use the same approaches you used back in the logistic regression class.)

Data Provided

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

- The training data set **insurance_t** contains 8,495 observations and selected 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 selected variables describing the customer's attributes **before** they were offered the new insurance product.
- The validation data set **insurance_v** contains 2,124 observations and selected variables.
- The table below describes the Roles and Description of the variables found in both data sets.
    - **Except for Branch of Bank**, consider anything with more than 10 distinct values as continuous.

Name	*Model Role*	*Description*
ACCTAGE	Input	Age of oldest account
DDA	Input	Indicator for checking account
DDABAL	Input	Checking account balance
DEP	Input	Checking deposits
DEPAMT	Input	Total amount deposited
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
INV	Input	Indicator for investment account
INVBAL	Input	INV balance
MM	Input	Indicator for money market account
MMBAL	Input	MM balance
MMCRED	Input	Number of money market credits
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
LORES	Input	Length of residence in years
HMVAL	Input	Value of home
AGE	Input	Age
CRSCORE	Input	Credit score
INAREA	Input	Indicator for local address
INS	Target	Indicator for purchase of insurance product
BRANCH	Input	Branch of bank