



# COMMERCIAL BANKING, CORP

COMMERCIAL BANKING, CORP

REQUEST FOR PROPOSAL

RFP #: IP – F3.H2

TITLE: BANKING INSURANCE PRODUCT – PHASE 2

CLOSING DATE AND TIME: NOVEMBER 19, 2021 @ 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.

<i>Name</i>	<i>Model Role</i>	<i>Description</i>
<i>ACCTAGE</i>	Input	Age of oldest account
<i>DDA</i>	Input	Indicator for checking account
<i>DDABAL</i>	Input	Checking account balance
<i>DEP</i>	Input	Checking deposits
<i>DEPAMT</i>	Input	Total amount deposited
<i>CHECKS</i>	Input	Number of checks written
<i>DIRDEP</i>	Input	Indicator for direct deposit
<i>NSF</i>	Input	Number of insufficient fund issues
<i>NSFAMT</i>	Input	Amount of NSF
<i>PHONE</i>	Input	Number of telephone banking interactions
<i>TELLER</i>	Input	Number of teller visit interactions
<i>SAV</i>	Input	Indicator for savings account
<i>SAVBAL</i>	Input	Savings account balance
<i>ATM</i>	Input	Indicator for ATM interaction
<i>ATMAMT</i>	Input	Total ATM withdrawal amount
<i>POS</i>	Input	Number of point of sale interactions
<i>POSAMT</i>	Input	Total amount for point of sale interactions
<i>CD</i>	Input	Indicator for certificate of deposit account
<i>CDBAL</i>	Input	CD balance
<i>IRA</i>	Input	Indicator for retirement account
<i>IRABAL</i>	Input	IRA balance
<i>INV</i>	Input	Indicator for investment account
<i>INVBAL</i>	Input	INV balance
<i>MM</i>	Input	Indicator for money market account
<i>MMBAL</i>	Input	MM balance
<i>MMCRED</i>	Input	Number of money market credits
<i>CC</i>	Input	Indicator for credit card
<i>CCBAL</i>	Input	CC balance
<i>CCPURC</i>	Input	Number of credit card purchases
<i>SDB</i>	Input	Indicator for safety deposit box
<i>INCOME</i>	Input	Income
<i>LORES</i>	Input	Length of residence in years
<i>HMVAL</i>	Input	Value of home
<i>AGE</i>	Input	Age
<i>CRSCORE</i>	Input	Credit score
<i>INAREA</i>	Input	Indicator for local address
<i>INS</i>	Target	Indicator for purchase of insurance product
<i>BRANCH</i>	Input	Branch of bank