Dustin Casey

5/4/17

MKTG 130

Professor Bryant

**Final Project Executive Summary**

Our team was given the important task of maximizing profit by creating a model to predict the likelihood that a prospect, when called, would do business with your bank as well as how much revenue would that sale/business provide. The solution to this problem comes in two parts: one being that we need to predict the probability of a sale; and, two being that we need to predict the potential revenue of the sale. To accomplish both of these objective, we’ve created regression models that analyze prospect data and characteristics (such as age, marital status, job, education, etc.) and create predictions to both of these questions. Before creating these models, we split our data up into a Training and Testing set to check the accuracy of our predictions as well as added a column called Purchase to both sets with either a 1 or a 0 based on if they had done business with your bank. We utilized multiple regression to determine which of these prospect characteristics where most important in determining potential revenue while using logistical regression to find the likelihood that a prospect would enter into business with you bank. After this, our team needed to decide who we should call based on their probability of purchase. We determined this probability threshold/cut-off-mark should be placed at 7.8% because it provided the highest true profit while limiting the amount of profitable calls missed. We then applied these models to test sets to confirm that they were accurate before applying it to a Validation list. We determined that calling prospects with a probability of purchase over 7.8% would produce the maximum profit for your bank. We reviewed these results by looking at the predicted revenue of the prospects we called minus the cost per call ($5) to get predicted profits.

**SAS CODE:**

**proc** **import** out=work.TrainData datafile="Z:\Students\100133984\MKTG 130\Final Project\Project Data - DCasey.xlsx" dbms=xlsx replace;

Sheet='Data';

**run**;

**proc** **import** out=work.ValidationData datafile="Z:\Students\100133984\MKTG 130\Final Project\Project Data - DCasey.xlsx" dbms=xlsx replace;

Sheet='Validation';

**run**;

**data** work.TrainData;

set work.TrainData;

if Gross\_Revenue = **0** then Purchase = **0**;

else Purchase=**1**;

**run**;

**data** work.RegTrainData;

set work.TrainData(where=(Purchase=**1**));

**run**;

**data** work.TestData;

set work.TrainData(where=(RandomID <= **10000**));

**run**;

**data** work.TrainData;

set work.TrainData(where=(RandomID > **10000**));

**run**;

ods graphics on;

**proc** **logistic** data= work.TrainData descending;

class job marital education default housing loan contact mon poutcome;

model Purchase = age job marital education default housing loan contact

mon days previous poutcome emp\_var\_rate cons\_price\_idx cons\_conf\_idx euribor3m nr\_employed / selection=stepwise ctable outroc=work.ROCData;

store work.LogisticModel;

**run**;

**proc** **glmselect** data=Work.RegTrainData plots=all ;

class job marital education default housing loan contact mon poutcome;

model Gross\_Revenue = age job marital education default housing loan contact

mon days previous poutcome emp\_var\_rate cons\_price\_idx cons\_conf\_idx euribor3m nr\_employed / selection=stepwise;

store work.RegressionModel;

**run**;

ods graphics off;

/\*Testing model on test set\*/

**proc** **plm** source=RegressionModel;

score data=work.TestData out=work.TestScored predicted = PredictedRevenue;

**run**;

**proc** **plm** source=LogisticModel;

score data=work.TestScored out=work.TestScored predicted = ProbabilityOfPurchase / ilink;

**run**;

**data** work.TestScored;

set work.TestScored;

label ProbabilityOfPurchase = "Probability of Purchase";

label PredictedRevenue = "Predicted Revenue";

if ProbabilityOfPurchase > **.0875** then PredictedPurchase = **1**;

else PredictedPurchase = **0**;

label PredictedPurchase = "Predicted Purchase";

if PredictedPurchase = **1** then PredictedProfit = (PredictedRevenue-**5**);

else PredictedProfit = **0**;

if PredictedPurchase = **1** and Purchase = **0** then PredictedProfit = -**5**;

label PredictedProfit = "Predicted Profit";

**run**;

**proc** **means** data=TestScored sum mean;

var PredictedProfit;

class PredictedPurchase;

**run**;

/\* Applying model to Validation Data Set\*/

**proc** **plm** source=RegressionModel;

score data=work.ValidationData out=work.ValidationScored predicted = PredictedRevenue;

**run**;

**proc** **plm** source=LogisticModel;

score data=work.ValidationScored out=work.ValidationScored predicted = ProbabilityOfPurchase / ilink;

**run**;

**data** work.ValidationScored;

set work.ValidationScored;

label ProbabilityOfPurchase = "Probability of Purchase";

label PredictedRevenue = "Predicted Revenue";

if ProbabilityOfPurchase > **.0875** then PredictedPurchase = **1**;

else PredictedPurchase = **0**;

label PredictedPurchase = "Predicted Purchase";

if PredictedPurchase = **1** then PredictedProfit = (PredictedRevenue-**5**);

if PredictedPurchase = **0** then PredictedProfit = **0**;

label PredictedProfit = "Predicted Profit";

**run**;

**proc** **means** data=ValidationScored sum mean;

var PredictedProfit;

class PredictedPurchase;

**run**;

/\*Tried out tree for logistic regression, but found more success with the proc logistic method.

ods graphics on;

proc hpsplit data=TrainData seed=1000 plots=all;

class Purchase job marital education default housing loan contact mon poutcome;

model Purchase (event='1') = age job marital education default housing loan contact

mon days previous poutcome emp\_var\_rate cons\_price\_idx cons\_conf\_idx euribor3m nr\_employed;

grow entropy;

prune costcomplexity;

ID RandomID;

output out=LogTreeOut;

run;

ods graphics off;

data TestTreeScored;

set TestData end=eof;

keep RandomNumber Actual Predicted Prob;

%include "Z:\Students\100133984\MKTG 130\Final Project\BuiltTreeCode.sas";

Actual = Purchase;

Prob = P\_Purchase1;

Predicted = (Prob >= 0.1);

run;

data LogValidateScored;

set ValidationData end=eof;

keep ID Predicted Prob;

%include "Z:\Students\100133984\MKTG 130\Final Project\BuiltTreeCode.sas";

Prob = P\_Purchase1;

Predicted = (Prob >= 0.1);

run;

\*/