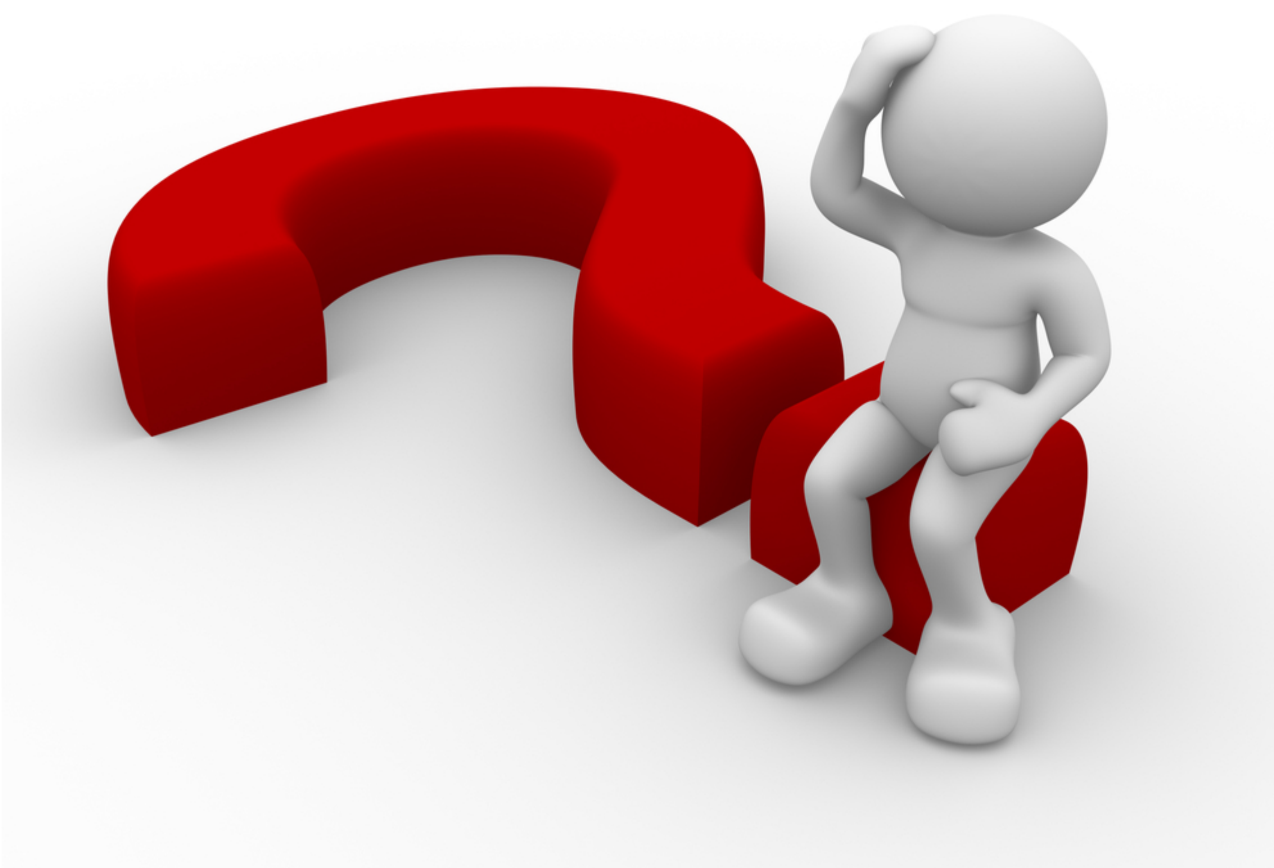
Title: Improving Hometown Banks Term Deposit Client Base By: Dylan Clark Date: April 25, 2017

The Question



Based off of the data set provided by Hometown Bank what are the major characteristics that Hometown Banks term deposit client base consist of? Additionally, is it possible to use these findings or explanatory variables in the data set to predict future customers into new deposit clients? If so how accurate would the model be?

Abstract: The following excerpt identifies the common characteristcs of Hometown Banks term deposit client base. It also provides insight to how accurate a model could be to help identify future term deposit customers of Hometown Bank from their existing and future client base. The excerpt provides a recommended model that Hometown Bank can adopt.

The Area of Opportunity & Data Set

Hometown Bank provided us a data set that provided a list of customers that consisted of customers that either had a term deposit or not. The data set that was provided contained 45,211 transactions that contained 16 explanatory variables that provided information from customer age, customer education leve, contact information, and various other varaibles that contained status of other demographics and campaign success ratings. The proble that Hometown Bank ran into was that they did not have a way to clean way to read and analyze the data, while being able to provide clear results of their direct marketing campaign to identify to identify new term deposit custoemrs. Thus, they turned to us to be able to read the data and analyze it for answers on questions they had.

Hometown Bank had the following questions: 1. What are the common characteristics of our customers that have term deposits? 2. Is it possible to identify current customers that should be targeted to buy term deposits with a model? 3. If so what would the accuracy of the model be? 4. What are some recommendations that can be executed to increase the term deposits customer base?

After given the set of questions from Hometow nBank I started to brainstorm about possible data cleansing activities and data mining techniques that could be used to clean the data and answer the first question. After brainstorming I knew that I could use R to help clean the data as well as use it to find out the characterisitcs of the customers that have term deposits. Additionally, I knew that could use the decision tree, bootstrapping and knn model to help create a useful model to help identify and predict current custoemrs that should be targeted to buy term deposites. From there I moved on to begin my plan of attack!

My Plan of Attack

I started with using R to begin my data cleaning. I loaded the appropriate information that I needed into R. Using R I started using some activities to load in certain external programs into the system to help identify the customer population that had short term deposits as well as restructure the data set to make the y..category or target variable appear first in the overall data set. Doing this makes it easier to keep track of as well as makes it easier to follow the programming created to explore the data and create a model to identify customers that may purchase a term depoit.

EDA - DATA PREPARATION

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.3.3
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```r
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.3.3
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 3.3.3
```

```r
library(adabag)
```

```
## Warning: package 'adabag' was built under R version 3.3.3
```

```
## Loading required package: rpart
```

```r
library(ROCR)
```

```
## Warning: package 'ROCR' was built under R version 3.3.3
```

```
## Loading required package: gplots
```

```
## Warning: package 'gplots' was built under R version 3.3.3
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(doSNOW)
```

```
## Warning: package 'doSNOW' was built under R version 3.3.3
```

```
## Loading required package: foreach
```

```
## Warning: package 'foreach' was built under R version 3.3.3
```

```
## Loading required package: iterators
```

```
## Loading required package: snow
```

```
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.3.3
```

```
setwd("C:/Users/dylan/Desktop/MBA/ADMSpring/Week_5")

BANK_FULL <- read.csv("bank_full.csv")
#reading in data set

BANK_FULL <- BANK_FULL[,c(17,1:16)]
#rearrandign  the columns to make the target variable to be the first

dim(BANK_FULL)
```

```
## [1] 45211    17
```

After cleaning the data I was able to begin my basic exploration to find out some key elements that I needed to utlimately start getting to the meat of some of the major questions that were asked by Hometown Bank. I started analyzing the data by using R to find out it’s data structure elements as well as limit down the data set to identify what the current term deposit client base characteristics were. I was able to run some basic distinct table counts to help identify the overall populations as well as to look at the characteristics trends of Hometown Banks current term deposit cleint base.

EDA CONTINUED - DISCOVERING THE DATA SET

```
dim(BANK_FULL)
```

```
## [1] 45211    17
```

```
str(BANK_FULL)
```

```
## 'data.frame':    45211 obs. of  17 variables:
##  $ y..category        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1
1 ...
##  $ age..number        : int  58 44 33 47 33 35 28 42 58 43 ...
##  $ job..category      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10
3 2 12 5 5 3 6 10 ...
##  $ marital..category  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2
3 2 3 1 2 3 ...
##  $ education..category: Factor w/ 4 levels "primary","secondary",..: 3 2 2
4 4 3 3 3 1 2 ...
##  $ default..category  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 2 1
1 ...
##  $ balance..number    : int  2143 29 2 1506 1 231 447 2 121 593 ...
##  $ housing..category  : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2
2 ...
##  $ loan..category     : Factor w/ 2 levels "no","yes": 1 1 2 1 1 1 2 1 1
1 ...
##  $ contact..category  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3
3 3 3 3 3 3 3 ...
##  $ day..number        : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ month..category    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9
9 9 9 9 9 ...
##  $ duration..number   : int  261 151 76 92 198 139 217 380 50 55 ...
##  $ campaign..number   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays..number      : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ previous..number   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome..category : Factor w/ 4 levels "failure","other",..: 4 4 4 4 4
4 4 4 4 4 ...
```

```
table(BANK_FULL$y..category)
```

```
##
##    no   yes
## 39922  5289
```

```
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 3.3.3
```

```
## Loading required package: gsubfn
```

```
## Warning: package 'gsubfn' was built under R version 3.3.3
```

```
## Loading required package: proto
```

```
## Warning: package 'proto' was built under R version 3.3.3
```

```
## Loading required package: RSQLite
```

```
## Warning: package 'RSQLite' was built under R version 3.3.3
```

```
#BANKFULL_SQL <- BANK_FULL

#BANKFULL_SQL$YCATEGORY <- BANKFULL_SQL$y..category

#TERM_DEPOSIT_CUSTOMERS <- sqldf("SELECT * FROM BANKFULL_SQL WHERE YCATEGORY li
ke 'yes' ")

#table(TERM_DEPOSIT_CUSTOMERS$poutcome..category) #about a 2% success
#table(TERM_DEPOSIT_CUSTOMERS$age..number) #larger numbers towards 25-49 - but
target 21
#table(TERM_DEPOSIT_CUSTOMERS$job..category) #large population towards manageme
nt, blue collared, admin and tech
#table(TERM_DEPOSIT_CUSTOMERS$marital..category) #mostly married
#table(TERM_DEPOSIT_CUSTOMERS$education..category) #majority is secondary and g
raduate"tertiary" schooling
#table(TERM_DEPOSIT_CUSTOMERS$marital..category) #
#table(TERM_DEPOSIT_CUSTOMERS$default..category) #not defaulted on a loan
#table(TERM_DEPOSIT_CUSTOMERS$housing..category)
#table(TERM_DEPOSIT_CUSTOMERS$loan..category) #no personal loan
```

I was able to identify that about 11.7% (5,289 out of 45,000 customers) of all Hometown Banks clients have a term deposit. It also seemed that only about 2% of the customers became term deposit holders from the market campaign that Hometown Bank executed, which is relatively poor performance compared to a normal 5% which is a successful direct marketing campaign success rate. However, after identifying the count I then spread the customer base into two populations. One population of customers that had a term deposit and the other non having term deposits. From there I was able to idetnify the common trends of charactersistics of customers that had term deposits.

I was able to identify that the common customer that had a term deposit more than likely had the following characteristics: - was between the ages of 25-49 years old - had a job in one of the following sectors: Admin, Blue-Collar, Management, Retired, Technician, Student - Management, Admin and Technician had the largest amounts - the customer either had a secondary or tertiary education level - the customer did not have defaulted credit - the customer did not have a personal loan

After identifying these commonalities I then started limiting down the customer population that did not have a term deposit since they were the current opporunity target that Hometown Bank could get business from. Hometown seemed to be focused on capitalizing on their current customer base that did not have a term deposit. Hometown wanted to turn those non term deposit holders into term deposit holders.I believed that it would be important to limit their non term deposit customer base based off of like characteristics of those that had term deposits. Thus, we would limit the non term deposit customer population down to custoemrs that were, between the ages of 25-49 years of age, that worked in the admin, management, technician sector or were a student or retited, that had a secondary or tertirary education level, did not have defaulted credit and that did not have a person loan. T

Limiting Down the Non Term Deposit Customer Population

```
#NON_TERM_DEPOSIT_CUSTOMERS <- sqldf("SELECT * FROM BANKFULL_SQL WHERE YCATEGORY like 'no' ")

#colnames(NON_TERM_DEPOSIT_CUSTOMERS_V3)[9] <- "personal" #used to rename columns in order to get sqldf to work and not error out on syntax problems
#NON_TERM_DEPOSIT_CUSTOMERS_V1 <- sqldf("SELECT * FROM NON_TERM_DEPOSIT_CUSTOMERS WHERE AGE BETWEEN 21 and 49 ")
#NON_TERM_DEPOSIT_CUSTOMERS_V2 <- sqldf("SELECT * FROM NON_TERM_DEPOSIT_CUSTOMERS_V1 WHERE job_category in #('admin.','blue-collar','management','retired','technician','student')  ")
#NON_TERM_DEPOSIT_CUSTOMERS_V3 <- sqldf("SELECT * FROM NON_TERM_DEPOSIT_CUSTOMERS_V3 WHERE personal in ('no') ")

#dim(NON_TERM_DEPOSIT_CUSTOMERS_V3)
```

After limiting down the non-term population we were able to identify a customer target population that Hometown Bank should really focus their direct marketing efforts on. The non term deposit holder population was taken down from approximately 40,000 customers down to approximately 16,000. This would help Hometown Bank to focus on target population of customers to send future marketing ads geared towards term deposits to. While it is good that Hometown Bank is able to idenify a common customer population they are ultimately also wanting to be able to know if they can develop an algorithm to determine future customers term deposit holding status as well as what the accuracy of what it would be. To help answer this question we turned to working with three seperate models to help determine a model that could work.

We used the following methods to help determine it: - Decision Tree - Bootstrapping - KNN

Below we analyze the work that was used to build the models and the outcomes of them. Before we conducted the models we first needed to prepare out data and models. We split our data into two sets using the 80/20 split method. The first set contained 80% of our overall observations and was going to be used as our training group that ultimately allow us to determine outcomes that could be valdiated against the second group of data that was our second group which was our validation group. The splitting of the data is conducted below:

Split the training sets and checking splits variances of target variable

```
set.seed(123)
BANK_RAND <- BANK_FULL[order(runif(45000)), ]

BANK_TRAIN <- BANK_RAND[1:36000, ] #Training data set; 36000 observations

BANK_VALIDATE <- BANK_RAND[36001:45000, ]

dim(BANK_TRAIN) #checking the split
```

```
## [1] 36000    17
```

```
dim(BANK_VALIDATE) #checking the split
```

```
## [1] 9000   17
```

```
prop.table(table(BANK_TRAIN$y..category))
```

```
##
##        no       yes
## 0.8846111 0.1153889
```

```
prop.table(table(BANK_VALIDATE$y..category))
```

```
##
##        no       yes
## 0.8851111 0.1148889
```

```
#11.5% versus 11.4%
```

MODELS USED FOR ANALYSIS

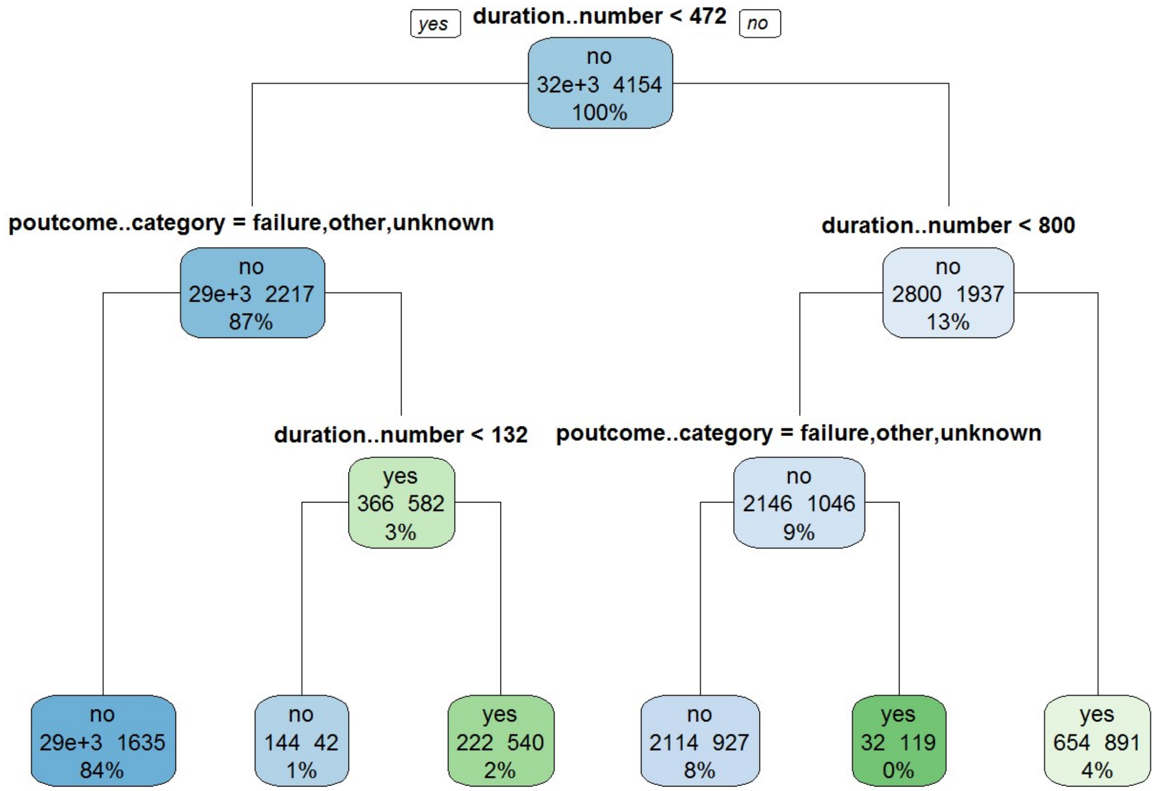
Before conducting our first model which is the decision tree method we wanted to ensure that there was equal representation of people that were term depsoit customers in both data sets. We were able to verify that there was even distribution of around 11.4% in both our train and validation data groups. Thus, we were able to validate that there is even distrubtion in the groups and that we were good to conduct our first method, the decision tree method.

MODEL 1 - DECISION TREE

```
set.seed(123)

BANK_RPART <- rpart(BANK_TRAIN$y..category~., method = "class", parms = list(sp
lit="gini"), data = BANK_TRAIN)

rpart.plot(BANK_RPART, type=1, extra=101)
```



The decision tree method simply worked to identify the most impactful variables and split the training data set into various groups based off of tree like path. The model determined that customers that were non term deposit holders had not been contacted for more than 472 seconds in their last interaction and had a failure, other or unknown cateogry outcome in the previous marketing campaign. Therefore, it appeared that close to more 80% of the customers were not given the attention in their last customer interaction wasn't sufficient. However, it appeared that customers that had duration contact time larger than 472 seconds had a larger success rate. Now that we were able to create the breakdown structure it is important to train and predict our findings based off of the decision tree model.

CARETING and ASSESS USABILITY OF MODEL

```
library(caret)
actual <- BANK_VALIDATE$y..category
predicted <- predict(BANK_RPART, BANK_VALIDATE, type="class")
results.matrix <- confusionMatrix(predicted, actual, positive="yes")
print(results.matrix)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  7733  678
##        yes  233  356
##
##                Accuracy : 0.8988
##                  95% CI : (0.8924, 0.9049)
##     No Information Rate : 0.8851
##     P-Value [Acc > NIR] : 1.908e-05
##
##                   Kappa : 0.3876
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.34429
##             Specificity : 0.97075
##          Pos Pred Value : 0.60441
##          Neg Pred Value : 0.91939
##              Prevalence : 0.11489
##          Detection Rate : 0.03956
##    Detection Prevalence : 0.06544
##       Balanced Accuracy : 0.65752
##
##        'Positive' Class : yes
##
```

Based off the decision tree model the accuracy of the predictions was accurate apprximately 90% of the time. So, the decision tree does a good job at being accurate. But the model has some work to do when classifying term holders with a low sensitivity raitn gof 34%, but it is really good when identifying non term holders with a 97% specifify rate. Thus, the method is overall fairly accurate, but highly accurate when classifying non term holders. So, the model appears to be helpful at times. But we should do some more analysis to help assess if there is a better model that could be used.

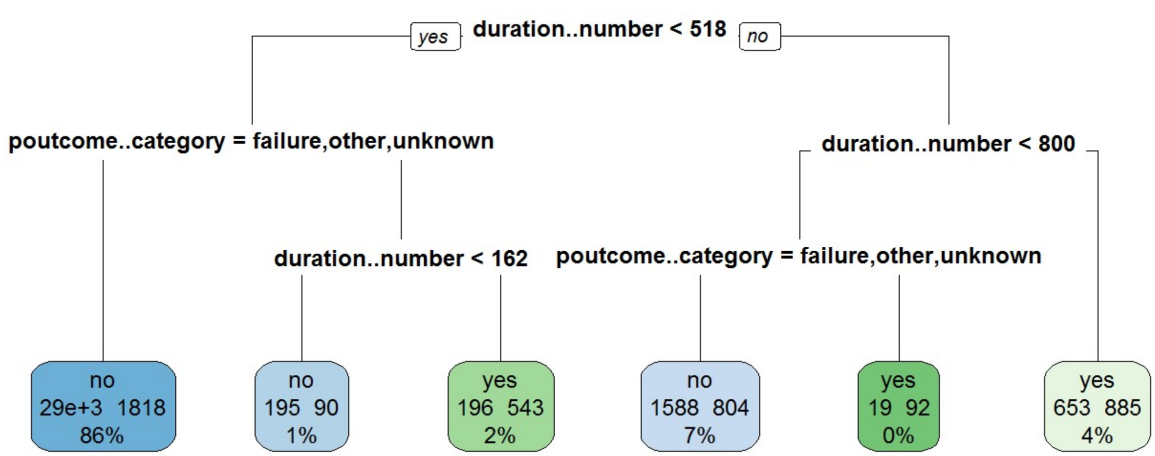
To help better assess the decision tree model we turn to the caret model. approach/

USING THE CARET PACKAGE TRAIN & VALIDATE MODELS

```
library(rpart)
set.seed(123)
trainIndex <- createDataPartition(BANK_FULL$y..category, p = .8, list = FALSE,t
imes = 1)
BANK_TRAIN_CARET <- BANK_FULL[ trainIndex,]
BANK_VALIDATE_CARET <- BANK_FULL [ -trainIndex,]

BANK_RPART_CARET <- rpart(y..category~., method = "class", parms = list(split
="gini"), data = BANK_TRAIN_CARET)

rpart.plot(BANK_RPART_CARET, type=0, extra=101)
```




```
CARET_ACTUAL <- BANK_VALIDATE_CARET$y..category
CARET_PREDICTED <- predict(BANK_RPART_CARET, BANK_VALIDATE_CARET, type="class")
CARET_results.matrix <- confusionMatrix(CARET_PREDICTED, CARET_ACTUAL, positive
="yes")
print(CARET_results.matrix)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  7755  672
##        yes  229  385
##
##                Accuracy : 0.9003
##                  95% CI : (0.894, 0.9064)
##     No Information Rate : 0.8831
##     P-Value [Acc > NIR] : 9.899e-08
##
##                   Kappa : 0.4101
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.36424
##             Specificity : 0.97132
##          Pos Pred Value : 0.62704
##          Neg Pred Value : 0.92026
##              Prevalence : 0.11691
##          Detection Rate : 0.04258
##    Detection Prevalence : 0.06791
##       Balanced Accuracy : 0.66778
##
##        'Positive' Class : yes
##
```

Accuracy=90% Sensitivity=36% Specificity=97%

The caret approach createsa better test set that shows the decision tree is doing slightly better at identifying term deposit holders. So, thus far the caret approach appears to be the best option when susing the decision tree method. However, we should try to use various methods to assess the a method performance.

Thus, we turn to bootstrapping which is a resampling technique that obtains distinct datasets by repeatedly gatehring sample observations from the original data set with replacements. Bootrapping is used by sampling with replacement and is the same size as the original data set. The bootstapp resampling technique is just a restructuring method to the decision tree method.

MODEL 2 - BOOTSTRAPPING MODEL

```
BOOTSTRAPCNTRL <- trainControl(method="boot", number=10)
set.seed(123)
BANK_BOOTSTRAP <- train(y..category~., data=BANK_TRAIN_CARET, method="rpart", m
etric="Accuracy", trControl=BOOTSTRAPCNTRL)
BANK_BOOTSTRAP
```

```
## CART
##
## 36170 samples
##    16 predictor
##     2 classes: 'no', 'yes'
##
## No pre-processing
## Resampling: Bootstrapped (10 reps)
## Summary of sample sizes: 36170, 36170, 36170, 36170, 36170, 36170, ...
## Resampling results across tuning parameters:
##
##   cp          Accuracy   Kappa
##   0.01724953  0.8995207  0.3835538
##   0.02481096  0.8973167  0.3792554
##   0.03733459  0.8909524  0.2254005
##
## Accuracy was used to select the optimal model using  the largest value.
## The final value used for the model was cp = 0.01724953.
```

```
ACTUAL_BOOTSTRAP <- BANK_VALIDATE_CARET$y..category
PREDITED_BOOTSTRAP <- predict(BANK_BOOTSTRAP, BANK_VALIDATE_CARET, type="raw")
BOOTSTRAP.results.matrix <- confusionMatrix(PREDITED_BOOTSTRAP, ACTUAL_BOOTSTRA
P, positive="yes")
print(BOOTSTRAP.results.matrix)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   no  yes
##        no  7763  691
##        yes  221  366
##
##                Accuracy : 0.8991
##                  95% CI : (0.8927, 0.9053)
##     No Information Rate : 0.8831
##     P-Value [Acc > NIR] : 7.004e-07
##
##                   Kappa : 0.3947
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.34626
##             Specificity : 0.97232
##          Pos Pred Value : 0.62351
##          Neg Pred Value : 0.91826
##              Prevalence : 0.11691
##          Detection Rate : 0.04048
##    Detection Prevalence : 0.06493
##       Balanced Accuracy : 0.65929
##
##        'Positive' Class : yes
##
```

Accuracy=90% Sensitivity=32% Specificity=97%

However, after running the bootstapping technique the model appears to be a bit less sensitive compared to the caret approach. However, it appears competitive in accuracy and specificy, but is not sensitive enough compared to the caret approch. Thus, the decision tree method with the caret approach still appears to be best model thus far. However, we will turn to the last method which is a KNN approach to better assess caret approach model performance.

The KNN method is much like a decision in the form that it is used to answer a yes/noquestion of a target variable. However, it differs by using the distance function to measure the performance and predicted outcomes. Much like the decision tree we used teh split the data up into sets. But instead of being train and validate the two groups are named train and test. The model will have 80% of the observations in the train data set and 20% in the test data set. The train group will be used to train our KNN model and then be used to predict the observations in the test set, much like the validate group in the decision tree method. We set k to 15 since it is believed based off previous tests that k = 15 provides the highest accuracy rate for the data set. From there we run our model to see the performance.

MODEL 3 - KNN MODEL

```
BANK_FULL_KNN <- BANK_FULL

BANK_FULL_KNN$y..category <-as.numeric(BANK_FULL_KNN$y..category)


BANK_N_TRAIN<- BANK_FULL_KNN[1:36000, 2:17]
BANK_N_TEST<- BANK_FULL_KNN[36001:45000, 2:17]

BANK_N_TRAIN_LABELS <- BANK_FULL_KNN[1:36000, 1]
BANK_N_TEST_LABELS <- BANK_FULL_KNN[36001:45000, 1]

library(class)
```

```
## Warning: package 'class' was built under R version 3.3.3
```

```
set.seed(123)

#BANK_PRED_KNN <- knn(train=BANK_N_TRAIN, test=BANK_N_TEST, cl=BANK_N_TRAIN_LAB
ELS, k=15)

library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 3.3.3
```

```r
#CrossTable(x=BANK_N_TEST_LABELS, y=BANK_PRED_KNN, prob.chisq=FALSE)

TP=144
TN=6162
FP=76
FN=2681


SENSITIVITY = TP/(TP+FN)
SPECIFICITY = TN/(TN+FP)
PRECISION = TP/(TP+FP)
ACCURACY = (TP+TN)/(TP+TN+FP+FN)

VALUE <- round(c(TP, TN, FP, FN, SENSITIVITY, SPECIFICITY, PRECISION, ACCURAC
Y), digits = 3)

Measure<-c("True Positive","True Negative","False Positive","False Negative","S
ensitivity/Recall=TP/(TN+FP)",
         "Specificity=TN/(TN+TP)","Precision=TP/(TP+FP)","Accuracy=(TP+TN)/tota
l")

table <- as.data.frame(cbind(Measure, VALUE))

library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.3.3
```

```r
kable(table)
```

Measure	VALUE
True Positive	144
True Negative	6162
False Positive	76
False Negative	2681
Sensitivity/Recall=TP/(TN+FP)	0.051
Specificity=TN/(TN+TP)	0.988
Precision=TP/(TP+FP)	0.655
Accuracy=(TP+TN)/total	0.696

Accuracy=70%

Measure	VALUE
Sensitivity=3%	
Specificity=99%	

Based off of the models performance the model is not accurate and is not specific enough in comparison to the caret approach using the decision tree method. Therfore, the decision tree method with the caret approach appears to be the best model to follow when trying to predict customers that may be term deposits holders. So, the answer to the second and third question is that yes a model can be used to predict and identify potential future term deposit holders in the current Hometown Bank customer base. The model that should be used based off of the models that were used in the analysis is the decision tree method with the caret approach, because it is the most accurate and does the best job at identifying term deposit holders and is very close to being the most specific at identifying non term deposit holders. The findings that the model provides will be used in the recommendations piece of the excerpt.

Recommendations

Based off the findings from the analysis Hometown Bank should do the following: - Refocus their term deposit product direct marketing efforts towards a specific target customer market.

These customers should have the following charactersistics: - be between the ages of 25-49 years old - have a job in one of the following sectors: Admin, Blue-Collar, Management, Retired, Technician, Student - the customer should have a secondary or tertiary education level - the customer should not have defaulted credit - the customer should not have a personal loan

Hometown bank should be more focused on developing a more targeted approach rather than trying to find a one size fits all approach in their marketing material. Trying to create more specific marketing material may be more successful if marketing material includes pictures or material that is more geared towards customers that have higher education levels, between the said ages and have a occupation in one of the specified industries. In the marketing material, Hometown Bank may offer a reduction in service fees on their customers accounts if the customer has a term deposit. This will provide an incentive to customers to sign up for a term deposit. Thus, not only does the bank sell more services, but in return the bank increases it’s financial assests, while saving the customers more money.

Additionally, Hometown Bank should encourage their personal bankers and tellers to spend more time developing relationships with their customers. From the decision tree model it was found that customers that had a term deposit had spent more time in their last encounter with the bank compared to those than did not have a term deposit. Thus, the term deposit customers appeared to spend more time at the branch. In fact, customers that had spent 8 minutes in duration in their last interaction had a more significant correlation to having a term deposit than those that did not have a term deposit. This may be simply because the time spent talking to the banker developed a relationship with a customer. Thus, when a teller or personal banker developed a relationship with the customer the customer appeared to be more open to recommendations on other services that Hometown has to offer. Therefore, Hometown Bank should either provide their frontline support (tellers & personal bankers) a financial incentive to increasing their average time spent with customers or should provide commission for each term deposit opened. If Hometown Bank does not want to provide a financiall incentive they could simply require their tellers and personal bankers to meet a said goal by a certain timeframe. Hometown could create a flagging tool on their customer operation account system that points out to the frontline support that the customer has a risk of being a term deposit customer based off of their characteristics. The flagging tool would provide the frontline support an easy way of being able to identify which customers should be spent more time with as well as which customer are more likely to become a term deposit holder. These are just some of the recommendations that could be taken into account from the fronline.

In conclusion, if Hometown Bank wishes to increase term deposit holder customer base they should, refocus their direct marketing efforts towards the specified target market, offer a customer incentive to having a term deposit, create a monitoring tool in thier customer operation system that identieis potential holders based off of findings in this analysis and the algorithm and provide a goal or financiall incentive to their frontline support for opening new term deposits.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.