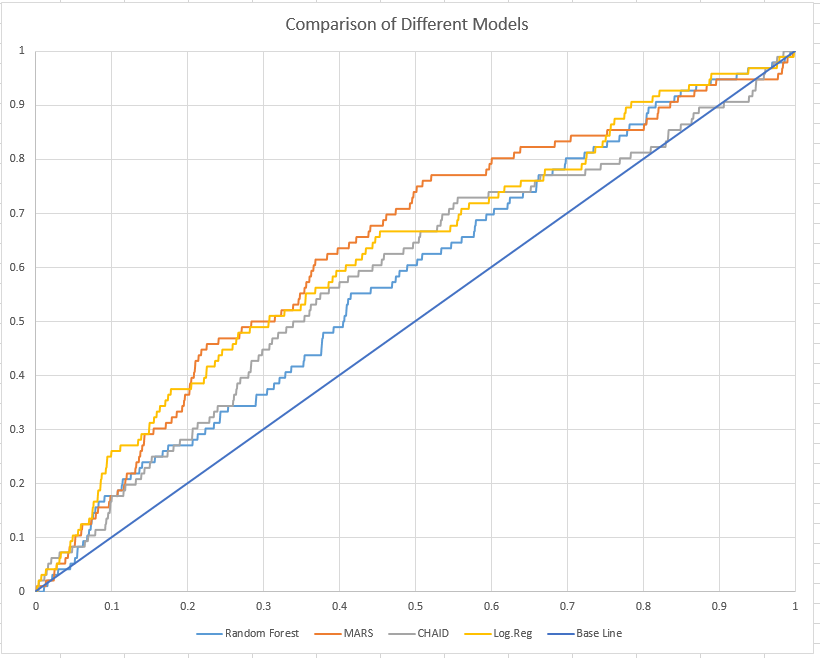
**GAIN CHART**

****

**Insights**

1. For 8% of the population, both logistic regression and Mars gives good prediction.
2. Between 8% to 20% of the population, Logistic regression gives better output.
3. Mars gives better gain and prediction for population between 20% to 75% and Logistic regression gives good prediction for final 25% of the population.
4. Both CHAID and Random Forest give poor performance.

**SAS** **Part 1**

**PROC** **IMPORT** OUT= WORK.coil

DATAFILE= "C:\Users\akalia3\Downloads\S13.csv"

DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

**data** coil2;

set coil ;

rand=ranuni(**092765**);

if rand <=**.7** then RespHoldout=**.**;

else if rand >**.7** then do;

RespHoldout=Resp;

Resp=**.**

;

end;

**run**;

**data** coil3;

set coil2;

array orig[**11**](**0**, **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, **9**, **10**);

array new[**11**] (**0**,**25**,**75**,**150**,**350**,**750**,**3000**,**7500**,**15000**,**30000**,**30000**);

retain orig1-orig10 new1-new10;

do i=**1** to dim(orig);

if pwapar=orig[i] then pwapar2=new[i];

if PAANHA=orig[i] then PAANHA2=new[i];

if PPERSA=orig[i] then PPERSA2=new[i];

end;

drop orig1--orig11 new1--new11 i;

**run**;

**proc** **freq** data=coil3;

tables pwapar\*pwapar2

PAANHA\*PAANHA2

PPERSA\*PPERSA2/list;

**run**;

**data** coil3;

set coil3;

drop pwapar paanha ppersa;

**run**;

**data** coil4;

set coil3;

array orig[**11**](**0**, **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, **9**, **10**);

array new[**11**] (**0**,**5.5**,**17**,**30**,**43**,**56**,**69**,**82**,**94**,**100**,**100**);

retain orig1-orig10 new1-new10;

do i=**1** to dim(orig);

if MGODRK =orig[i] then MGODRK2 =new[i];

if MGODPR =orig[i] then MGODPR2 =new[i];

if MRELGE =orig[i] then MRELGE2 =new[i];

if MFALLE =orig[i] then MFALLE2 =new[i];

if MFWEKI =orig[i] then MFWEKI2 =new[i];

if MOPLHO =orig[i] then MOPLHO2 =new[i];

if MSKA =orig[i] then MSKA2 =new[i];

if MSKB1 =orig[i] then MSKB12 =new[i];

if MSKB2 =orig[i] then MSKB22 =new[i];

if MSKC =orig[i] then MSKC2 =new[i];

if MHHUUR =orig[i] then MHHUUR2 =new[i];

if MAUT1 =orig[i] then MAUT12 =new[i];

if MAUT2 =orig[i] then MAUT22 =new[i];

if MAUT0 =orig[i] then MAUT02 =new[i];

if MINKGE =orig[i] then MINKGE2 =new[i];

end;

drop orig1--orig11 new1--new11 i;

**run**;

**proc** **freq** data=coil4;

tables

MGODRK\*MGODRK2

MGODPR\*MGODPR2

MRELGE\*MRELGE2

MFALLE\*MFALLE2

MFWEKI\*MFWEKI2

MOPLHO\*MOPLHO2

MSKA\*MSKA2

MSKB1\*MSKB12

MSKB2\*MSKB22

MSKC\*MSKC2

MHHUUR\*MHHUUR2

MAUT1\*MAUT12

MAUT2\*MAUT22

MAUT0\*MAUT02

MINKGE\*MINKGE2

/list;

**run**;

**data** coil5;

set coil4;

drop

MGODRK

MGODPR

MRELGE

MFALLE

MFWEKI

MOPLHO

MSKA

MSKB1

MSKB2

MSKC

MHHUUR

MAUT1

MAUT2

MAUT0

MINKGE

;

**run**;

%***CatToBinWithDrop***(coil5,seqnum,mostyp);

%***CatToBinWithDrop***(coil5,seqnum,MOSHOO);

**proc** **means** data=coil5 n nmiss;

**run**;

**data** hold00;

set coil5;

if resp=**.**;

**run**;

**data** anal00;

set coil5;

if resp>**.**;

**run**;

**PROC** **EXPORT** DATA= WORK.ANAL00

OUTFILE= "C:\Users\akalia3\Desktop\Anal00.csv"

DBMS=CSV REPLACE;

PUTNAMES=YES;

**RUN**;

**PROC** **EXPORT** DATA= WORK.Hold00

OUTFILE= "C:\Users\akalia3\Desktop\Hold00.csv"

DBMS=CSV REPLACE;

PUTNAMES=YES;

**RUN**;

**proc** **contents** data=anal00;

**run**;

**R Code**

setwd("C:/Users/Abilash/Desktop/Final")

train<-read.csv("Anal00.csv")

train<-subset(train,select=-c(rand,RespHoldout))

# implementing Random Forest model

library(randomForest)

rf<-randomForest(x=train[-c(1,8)],y=train$Resp,ntree=200,importance=TRUE)

varImpPlot(rf)

test<-read.csv("Hold00.csv")

test<-subset(test,select=-c(Resp))

test$Resp<-test$RespHoldout

test<-subset(test,select=-c(rand,RespHoldout))

test$rf<-predict(rf,test)

# installing "earth" package

install.packages("earth")

library (earth)

# implementing mars model

mars<-earth(x=train[-c(1,8)],y=train$Resp)

summary(mars,digits=2,style="pmax")

test$marspred<-predict(mars,test)

summary(mars,digits=2,style="pmax")

# implementing CHAID model

tree<-rpart(Resp~., data=train[-c(1)], control=rpart.control(cp=.005))

plot(tree, uniform=TRUE,main="Classification Tree for COIL Data")

text(tree, use.n=TRUE, all=TRUE, cex=.8)

test$treepred<-predict(tree,test)

Call: earth(x=train[-c(1,8)],

y=train$Resp)

train$Resp =

0.21

+ 0.12 \* mostyp\_4

+ 0.051 \* mostyp\_28

+ 0.063 \* mostyp\_29

+ 0.071 \* mostyp\_36

+ 0.028 \* MOSHOO\_3

+ 0.056 \* MOSHOO\_6

+ 0.058 \* pmax(0, AWAPAR - 1)

- 0.0024 \* pmax(0, 75 - pwapar2)

- 0.00057 \* pmax(0, pwapar2 - 75)

+ 0.0011 \* pmax(0, 30 - MRELGE2)

+ 0.0027 \* pmax(0, MFWEKI2 - 56)

- 0.024 \* pmax(0, MFWEKI2 - 94)

- 0.004 \* pmax(0, 5.5 - MSKA2)

+ 0.00062 \* pmax(0, 43 - MINKGE2)

**SAS Part 2**

**PROC** **IMPORT** OUT= WORK.FromR

DATAFILE= "C:\Users\akalia3\Desktop\test.csv"

DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

**proc** **freq** data=FromR;

tables resp;

**run**;

\*If this doesnt match your mid-term memo then contact me. SAS read in your file wrong;

**proc** **means** data=work.FromR nmiss mean std cv p1 p10 p25 p50 p75 p90 p99;

var treepred marspred rf;

**run**;

**data** treeanal;

set FromR;

keep treepred resp;

**run**;

**proc** **sort** data=treeanal;

by descending treepred;

**run**;

**data** treeanal2;

set treeanal;

treecumresp+resp;

treepct=treecumresp/**96**;

**run**;

**data** Forestanal;

set FromR;

keep rf resp;

**run**;

**proc** **sort** data=Forestanal;

by descending rf;

**run**;

**data** Forestanal2;

set Forestanal;

Forestcumresp+resp;

Forestpct=Forestcumresp/**96**;

**run**;

**data** Marsanal;

set FromR;

keep marspred resp;

**run**;

**proc** **sort** data=Marsanal;

by descending marspred;

**run**;

**data** Marsanal2;

set Marsanal;

Marscumresp+resp;

Marspct=Marscumresp/**96**;

**run**;

**data** compare;

merge ForestAnal2

MarsAnal2

TreeAnal2

RespAnal

;

**run**;