Task 2

**C5.0**

EL modelo utilizado para generar la predicción fue C5.0, ya que este presentó los mejores parámetros y una predicción certera de lo que se desea predecir.

SurveyData<-read.csv("CompleteResponses.csv",header=TRUE,sep=",")  
SurveyData$brand<-as.factor(SurveyData$brand)  
set.seed(456)  
SurveyData<-SurveyData[sample(1:nrow(SurveyData),7000,replace=FALSE),]  
index<-createDataPartition(SurveyData$brand,p=0.75,list=FALSE)  
trainSet<-SurveyData[index,]  
testSet<-SurveyData[-index,]  
vars<-c("salary","age")  
fitControl<-trainControl(method="repeatedcv",number=10,repeats=5)  
treeMod<-C5.0(x=trainSet[,vars],y=trainSet$brand,trControl=fitControl,tuneLenght=2)  
treeMod  
summary(treeMod)  
print(treeMod)

predicted<-matrix(rnorm(50),ncol=5)  
observed<-rnorm(10)  
apply(predicted,2,postResample,obs=observed)  
classes<-c("class1","class2")  
set.seed(1)

dat<data.frame(obs=factor(sample(classes,50,replace=TRUE)),pred=factor(sample(classes,50,replace=TRUE)),class1=runif(50))  
dat$class2<-1-dat$class1  
defaultSummary(dat,lev=classes)

varImp(object=treeMod)

InSurvey<-read.csv("SurveyIncomplete.csv",header=TRUE,sep=",")  
InSurvey$brand<-as.factor(InSurvey$brand)  
prediction<-predict(treeMod,InSurvey,type="class")  
prediction  
summary(prediction)

postResample(prediction,InSurvey$brand)

table(prediction,InSurvey$brand)

options(max.print=999999)

Evaluation on training data (7501 cases):

Decision Tree

----------------

Size Errors

12 1335(17.8%) <<

(a) (b) <-classified as

---- ----

2710 128 (a): class 0

1207 3456 (b): class 1

Attribute usage:

100.00% salary

80.63% age

> apply(predicted,2,postResample,obs=observed)

[,1] [,2] [,3] [,4] [,5]

RMSE 1.768918966 2.20435002 1.75570190 1.20829294 1.76940290

Rsquared 0.007577096 0.03329133 0.03677654 0.07521033 0.01860952

MAE 1.379716661 1.86780569 1.54019532 1.03895461 1.45841121

> varImp(object=treeMod)

Overall

salary 100.00

age 72.02

> summary(prediction)

0 1

2529 2471

> table(prediction,InSurvey$brand)

prediction 0

0 2529

1 2471

> postResample(prediction,InSurvey$brand)

Accuracy Kappa

1 NA