```
#Decision Tree Method 1
myd=read.table("census_small_tree_bal.csv", header = T, sep=',')
head(myd)
myd[myd==99]=NA
myd[myd==98]=NA
myd[myd==999]=NA
myd[myd==-1]=NA
myd$sex=factor(myd$sex, levels = c(1:2), labels = c("male","female"))
myd$wrkstat=factor(myd$wrkstat, levels = c(1:8), labels =
c("Fulltime","Part-
time","Tempnotwork","Unempl","Retired","School","Housekeep","Other"))
myd$happy=factor(myd$happy,levels = c(1:2), labels = c("Very
Happy","Not Happy"))
myd$marital=factor(myd$marital, levels = c(1:5), labels =
c("Married","Widowed","Divorced","Separated","Never Married"))
myd$polviews2=factor(myd$polviews2, levels = c(1:3), labels =
c("Libera","Moderate","Conservative"))
myd$satjob2=factor(myd$satjob2, levels = c(1:2), labels =
c("Satisfied","Not Satisfied"))
myd$agebin=cut(myd$age, c(18,30,45,65,90))
myd$postlife=factor(myd$postlife, levels = c(1:2), labels =
c("Yes","No"))
myd$health2=factor(myd$health2, levels=c(1:3), labels =
c("Excellent","Good","Poor"))
myd$wrkstat2=myd$wrkstat
myd$wrkstat2[myd$wrkstat2=="other"]="Tempnotwork"
myd$wrkstat2[myd$wrkstat2=="School"]="Tempnotwork"
myd$wrkstat2=factor(myd$wrkstat2, levels = c(1:5), labels =
c("Fulltime","Part-time","Tempnotwork","Unempl","Retired"))
myd$class=factor(myd$class, levels = c(1:4), labels =
c("Lower","Working","Middle","Upper"))
myd$born=factor(myd$born, levels = c(1:2),labels = c("Born in US","Not
Born in US"))
summary(myd[,-1])
listvar=c("happy","health2","agebin","satjob2","postlife","polviews2",
"marital","wrkstat2","sex","born","class")
myddt=myd[listvar]
train.size=round(0.8*nrow(myddt))
set.seed(1234)
id.train=sample(1:nrow(myddt), train.size, replace = FALSE)
myddt.train=myddt[id.train,]
mydtt.val=myddt[-id.train,]
library(rpart)
dtree=rpart(happy~., data=myddt.train, method="class",
parms=list(split="aini"))
barplot(dtree$variable.importance)
library(rpart.plot)
prp(dtree, type=4, extra=104, fallen.leaves = FALSE, main="Decision
Tree", faclen = 0)
print(dtree)
```

```
plotcp(dtree)
dtree.pred=predict(dtree, newdata = mydtt.val, type="class")
dtree.perf=table(mydtt.val$happy, dtree.pred,
dnn=c("Actual","Predicted"))
prop.table(dtree.perf,1)

#Decision Tree Method 2
myd=read.csv("winedata.csv", header=T)
head(myd)
library(rpart)
train.size=round(0.66*nrow(myd))
id.train=sample(1:nrow(myd), train.size, replace = FALSE)
myd.train=myd[id.train,]
myd.test=myd[-id.train,]
dtree=rpart(grape~., data=myd.train, method="class",
parms=list(split="gini"))
train.pred=predict(dtree, newdata=myd.train, type="class")
train.perf=table(myd.train$grape, train.pred,
dnn=c("Acutual","Predicted"))
prop.table(train.perf,1)
accuracy=sum(diag(train.perf))/sum(train.perf)
accuracy
```