Homework-4

**Chinki**

**April 29, 2017**

**Identifying risky bank loan.**

**Step 1: Collecting the data**

source of the data <http://archive.ics.uci.edu/ml.The> dataset contains information on loans obtained from a credit agency in Germany.The credit dataset includes 1,000 examples on loans, plus a set of numeric and nominal features indicating the characteristics of the loan and the loan applicant. A class variable indicates whether the loan went into default.

**Step 2: Exploring and preparing the data**

#Reading the data  
credit <- read.csv("credit.csv")

str(credit)

## 'data.frame': 1000 obs. of 17 variables:  
## $ checking\_balance : Factor w/ 4 levels "< 0 DM","> 200 DM",..: 1 3 4 1 1 4 4 3 4 3 ...  
## $ months\_loan\_duration: int 6 48 12 42 24 36 24 36 12 30 ...  
## $ credit\_history : Factor w/ 5 levels "critical","good",..: 1 2 1 2 4 2 2 2 2 1 ...  
## $ purpose : Factor w/ 6 levels "business","car",..: 5 5 4 5 2 4 5 2 5 2 ...  
## $ amount : int 1169 5951 2096 7882 4870 9055 2835 6948 3059 5234 ...  
## $ savings\_balance : Factor w/ 5 levels "< 100 DM","> 1000 DM",..: 5 1 1 1 1 5 4 1 2 1 ...  
## $ employment\_duration : Factor w/ 5 levels "< 1 year","> 7 years",..: 2 3 4 4 3 3 2 3 4 5 ...  
## $ percent\_of\_income : int 4 2 2 2 3 2 3 2 2 4 ...  
## $ years\_at\_residence : int 4 2 3 4 4 4 4 2 4 2 ...  
## $ age : int 67 22 49 45 53 35 53 35 61 28 ...  
## $ other\_credit : Factor w/ 3 levels "bank","none",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ housing : Factor w/ 3 levels "other","own",..: 2 2 2 1 1 1 2 3 2 2 ...  
## $ existing\_loans\_count: int 2 1 1 1 2 1 1 1 1 2 ...  
## $ job : Factor w/ 4 levels "management","skilled",..: 2 2 4 2 2 4 2 1 4 1 ...  
## $ dependents : int 1 1 2 2 2 2 1 1 1 1 ...  
## $ phone : Factor w/ 2 levels "no","yes": 2 1 1 1 1 2 1 2 1 1 ...  
## $ default : Factor w/ 2 levels "no","yes": 1 2 1 1 2 1 1 1 1 2 ...

#Looking for checking balance   
table(credit$checking\_balance)

##   
## < 0 DM > 200 DM 1 - 200 DM unknown   
## 274 63 269 394

#Looking for saving balance   
table(credit$savings\_balance)

##   
## < 100 DM > 1000 DM 100 - 500 DM 500 - 1000 DM unknown   
## 603 48 103 63 183

#summary of duration of loan  
summary(credit$months\_loan\_duration)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.0 12.0 18.0 20.9 24.0 72.0

#Summary of credit amont  
summary(credit$amount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 250 1366 2320 3271 3972 18420

#Looking for class variable credit default  
table(credit$default)

##   
## no yes   
## 700 300

#Creating a random sample   
set.seed(123)  
train\_sample <- sample(1000, 900)  
str(train\_sample)

## int [1:900] 288 788 409 881 937 46 525 887 548 453 ...

# Spliting the data into training and test data  
credit\_train <- credit[train\_sample, ]  
credit\_test <- credit[-train\_sample, ]

#checking default proportion in traning and test data  
prop.table(table(credit\_train$default))

##   
## no yes   
## 0.7033333 0.2966667

prop.table(table(credit\_test$default))

##   
## no yes   
## 0.67 0.33

**Step 3: Training a model on the data**

library(C50)  
credit\_model <- C5.0(credit\_train[-17], credit\_train$default)

#Credit\_model display  
credit\_model

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default)  
##   
## Classification Tree  
## Number of samples: 900   
## Number of predictors: 16   
##   
## Tree size: 57   
##   
## Non-standard options: attempt to group attributes

#Getting summary of credit\_model  
summary(credit\_model)

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon May 01 07:39:50 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 900 cases (17 attributes) from undefined.data  
##   
## Decision tree:  
##   
## checking\_balance in {> 200 DM,unknown}: no (412/50)  
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...credit\_history in {perfect,very good}: yes (59/18)  
## credit\_history in {critical,good,poor}:  
## :...months\_loan\_duration <= 22:  
## :...credit\_history = critical: no (72/14)  
## : credit\_history = poor:  
## : :...dependents > 1: no (5)  
## : : dependents <= 1:  
## : : :...years\_at\_residence <= 3: yes (4/1)  
## : : years\_at\_residence > 3: no (5/1)  
## : credit\_history = good:  
## : :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (15/1)  
## : savings\_balance = 100 - 500 DM:  
## : :...other\_credit = bank: yes (3)  
## : : other\_credit in {none,store}: no (9/2)  
## : savings\_balance = unknown:  
## : :...other\_credit = bank: yes (1)  
## : : other\_credit in {none,store}: no (21/8)  
## : savings\_balance = < 100 DM:  
## : :...purpose in {business,car0,renovations}: no (8/2)  
## : purpose = education:  
## : :...checking\_balance = < 0 DM: yes (4)  
## : : checking\_balance = 1 - 200 DM: no (1)  
## : purpose = car:  
## : :...employment\_duration = > 7 years: yes (5)  
## : : employment\_duration = unemployed: no (4/1)  
## : : employment\_duration = < 1 year:  
## : : :...years\_at\_residence <= 2: yes (5)  
## : : : years\_at\_residence > 2: no (3/1)  
## : : employment\_duration = 1 - 4 years:  
## : : :...years\_at\_residence <= 2: yes (2)  
## : : : years\_at\_residence > 2: no (6/1)  
## : : employment\_duration = 4 - 7 years:  
## : : :...amount <= 1680: yes (2)  
## : : amount > 1680: no (3)  
## : purpose = furniture/appliances:  
## : :...job in {management,unskilled}: no (23/3)  
## : job = unemployed: yes (1)  
## : job = skilled:  
## : :...months\_loan\_duration > 13: [S1]  
## : months\_loan\_duration <= 13:  
## : :...housing in {other,own}: no (23/4)  
## : housing = rent:  
## : :...percent\_of\_income <= 3: yes (3)  
## : percent\_of\_income > 3: no (2)  
## months\_loan\_duration > 22:  
## :...savings\_balance = > 1000 DM: no (2)  
## savings\_balance = 500 - 1000 DM: yes (4/1)  
## savings\_balance = 100 - 500 DM:  
## :...credit\_history in {critical,poor}: no (14/3)  
## : credit\_history = good:  
## : :...other\_credit = bank: no (1)  
## : other\_credit in {none,store}: yes (12/2)  
## savings\_balance = unknown:  
## :...checking\_balance = 1 - 200 DM: no (17)  
## : checking\_balance = < 0 DM:  
## : :...credit\_history = critical: no (1)  
## : credit\_history in {good,poor}: yes (12/3)  
## savings\_balance = < 100 DM:  
## :...months\_loan\_duration > 47: yes (21/2)  
## months\_loan\_duration <= 47:  
## :...housing = other:  
## :...percent\_of\_income <= 2: no (6)  
## : percent\_of\_income > 2: yes (9/3)  
## housing = rent:  
## :...other\_credit = bank: no (1)  
## : other\_credit in {none,store}: yes (16/3)  
## housing = own:  
## :...employment\_duration = > 7 years: no (13/4)  
## employment\_duration = 4 - 7 years:  
## :...job in {management,skilled,  
## : : unemployed}: yes (9/1)  
## : job = unskilled: no (1)  
## employment\_duration = unemployed:  
## :...years\_at\_residence <= 2: yes (4)  
## : years\_at\_residence > 2: no (3)  
## employment\_duration = 1 - 4 years:  
## :...purpose in {business,car0,education}: yes (7/1)  
## : purpose in {furniture/appliances,  
## : : renovations}: no (7)  
## : purpose = car:  
## : :...years\_at\_residence <= 3: yes (3)  
## : years\_at\_residence > 3: no (3)  
## employment\_duration = < 1 year:  
## :...years\_at\_residence > 3: yes (5)  
## years\_at\_residence <= 3:  
## :...other\_credit = bank: no (0)  
## other\_credit = store: yes (1)  
## other\_credit = none:  
## :...checking\_balance = 1 - 200 DM: no (8/2)  
## checking\_balance = < 0 DM:  
## :...job in {management,skilled,  
## : unemployed}: yes (2)  
## job = unskilled: no (3/1)  
##   
## SubTree [S1]  
##   
## employment\_duration in {< 1 year,4 - 7 years}: no (4)  
## employment\_duration in {> 7 years,1 - 4 years,unemployed}: yes (10)  
##   
##   
## Evaluation on training data (900 cases):  
##   
## Decision Tree   
## ----------------   
## Size Errors   
##   
## 56 133(14.8%) <<  
##   
##   
## (a) (b) <-classified as  
## ---- ----  
## 598 35 (a): class no  
## 98 169 (b): class yes  
##   
##   
## Attribute usage:  
##   
## 100.00% checking\_balance  
## 54.22% credit\_history  
## 47.67% months\_loan\_duration  
## 38.11% savings\_balance  
## 14.33% purpose  
## 14.33% housing  
## 12.56% employment\_duration  
## 9.00% job  
## 8.67% other\_credit  
## 6.33% years\_at\_residence  
## 2.22% percent\_of\_income  
## 1.56% dependents  
## 0.56% amount  
##   
##   
## Time: 0.0 secs

**Step 4: Evaluating model performance**

credit\_pred <- predict(credit\_model, credit\_test)  
library(gmodels)  
CrossTable(credit\_test$default, credit\_pred,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 100   
##   
##   
## | predicted default   
## actual default | no | yes | Row Total |   
## ---------------|-----------|-----------|-----------|  
## no | 59 | 8 | 67 |   
## | 0.590 | 0.080 | |   
## ---------------|-----------|-----------|-----------|  
## yes | 19 | 14 | 33 |   
## | 0.190 | 0.140 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 78 | 22 | 100 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Accuracy is 73% from this confusion matrix.

**Step 5: Improving model performance**

using boosting to incerase accuracy. using boosted decision tree with 10 trails.

credit\_boost10 <- C5.0(credit\_train[-17], credit\_train$default,  
 trials = 10)  
credit\_boost10

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default, trials = 10)  
##   
## Classification Tree  
## Number of samples: 900   
## Number of predictors: 16   
##   
## Number of boosting iterations: 10   
## Average tree size: 47.5   
##   
## Non-standard options: attempt to group attributes

#Summary of boosted decision tree  
summary(credit\_boost10)

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default, trials = 10)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon May 01 07:39:50 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 900 cases (17 attributes) from undefined.data  
##   
## ----- Trial 0: -----  
##   
## Decision tree:  
##   
## checking\_balance in {> 200 DM,unknown}: no (412/50)  
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...credit\_history in {perfect,very good}: yes (59/18)  
## credit\_history in {critical,good,poor}:  
## :...months\_loan\_duration <= 22:  
## :...credit\_history = critical: no (72/14)  
## : credit\_history = poor:  
## : :...dependents > 1: no (5)  
## : : dependents <= 1:  
## : : :...years\_at\_residence <= 3: yes (4/1)  
## : : years\_at\_residence > 3: no (5/1)  
## : credit\_history = good:  
## : :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (15/1)  
## : savings\_balance = 100 - 500 DM:  
## : :...other\_credit = bank: yes (3)  
## : : other\_credit in {none,store}: no (9/2)  
## : savings\_balance = unknown:  
## : :...other\_credit = bank: yes (1)  
## : : other\_credit in {none,store}: no (21/8)  
## : savings\_balance = < 100 DM:  
## : :...purpose in {business,car0,renovations}: no (8/2)  
## : purpose = education:  
## : :...checking\_balance = < 0 DM: yes (4)  
## : : checking\_balance = 1 - 200 DM: no (1)  
## : purpose = car:  
## : :...employment\_duration = > 7 years: yes (5)  
## : : employment\_duration = unemployed: no (4/1)  
## : : employment\_duration = < 1 year:  
## : : :...years\_at\_residence <= 2: yes (5)  
## : : : years\_at\_residence > 2: no (3/1)  
## : : employment\_duration = 1 - 4 years:  
## : : :...years\_at\_residence <= 2: yes (2)  
## : : : years\_at\_residence > 2: no (6/1)  
## : : employment\_duration = 4 - 7 years:  
## : : :...amount <= 1680: yes (2)  
## : : amount > 1680: no (3)  
## : purpose = furniture/appliances:  
## : :...job in {management,unskilled}: no (23/3)  
## : job = unemployed: yes (1)  
## : job = skilled:  
## : :...months\_loan\_duration > 13: [S1]  
## : months\_loan\_duration <= 13:  
## : :...housing in {other,own}: no (23/4)  
## : housing = rent:  
## : :...percent\_of\_income <= 3: yes (3)  
## : percent\_of\_income > 3: no (2)  
## months\_loan\_duration > 22:  
## :...savings\_balance = > 1000 DM: no (2)  
## savings\_balance = 500 - 1000 DM: yes (4/1)  
## savings\_balance = 100 - 500 DM:  
## :...credit\_history in {critical,poor}: no (14/3)  
## : credit\_history = good:  
## : :...other\_credit = bank: no (1)  
## : other\_credit in {none,store}: yes (12/2)  
## savings\_balance = unknown:  
## :...checking\_balance = 1 - 200 DM: no (17)  
## : checking\_balance = < 0 DM:  
## : :...credit\_history = critical: no (1)  
## : credit\_history in {good,poor}: yes (12/3)  
## savings\_balance = < 100 DM:  
## :...months\_loan\_duration > 47: yes (21/2)  
## months\_loan\_duration <= 47:  
## :...housing = other:  
## :...percent\_of\_income <= 2: no (6)  
## : percent\_of\_income > 2: yes (9/3)  
## housing = rent:  
## :...other\_credit = bank: no (1)  
## : other\_credit in {none,store}: yes (16/3)  
## housing = own:  
## :...employment\_duration = > 7 years: no (13/4)  
## employment\_duration = 4 - 7 years:  
## :...job in {management,skilled,  
## : : unemployed}: yes (9/1)  
## : job = unskilled: no (1)  
## employment\_duration = unemployed:  
## :...years\_at\_residence <= 2: yes (4)  
## : years\_at\_residence > 2: no (3)  
## employment\_duration = 1 - 4 years:  
## :...purpose in {business,car0,education}: yes (7/1)  
## : purpose in {furniture/appliances,  
## : : renovations}: no (7)  
## : purpose = car:  
## : :...years\_at\_residence <= 3: yes (3)  
## : years\_at\_residence > 3: no (3)  
## employment\_duration = < 1 year:  
## :...years\_at\_residence > 3: yes (5)  
## years\_at\_residence <= 3:  
## :...other\_credit = bank: no (0)  
## other\_credit = store: yes (1)  
## other\_credit = none:  
## :...checking\_balance = 1 - 200 DM: no (8/2)  
## checking\_balance = < 0 DM:  
## :...job in {management,skilled,  
## : unemployed}: yes (2)  
## job = unskilled: no (3/1)  
##   
## SubTree [S1]  
##   
## employment\_duration in {< 1 year,4 - 7 years}: no (4)  
## employment\_duration in {> 7 years,1 - 4 years,unemployed}: yes (10)  
##   
## ----- Trial 1: -----  
##   
## Decision tree:  
##   
## checking\_balance = unknown:  
## :...other\_credit in {bank,store}:  
## : :...purpose in {business,education,renovations}: yes (19.5/6.3)  
## : : purpose in {car0,furniture/appliances}: no (24.8/6.6)  
## : : purpose = car:  
## : : :...dependents <= 1: yes (20.1/4.8)  
## : : dependents > 1: no (2.4)  
## : other\_credit = none:  
## : :...credit\_history in {critical,perfect,very good}: no (102.8/4.4)  
## : credit\_history = good:  
## : :...existing\_loans\_count <= 1: no (112.7/17.5)  
## : : existing\_loans\_count > 1: yes (18.9/7.9)  
## : credit\_history = poor:  
## : :...years\_at\_residence <= 1: yes (4.4)  
## : years\_at\_residence > 1:  
## : :...percent\_of\_income <= 3: no (11.9)  
## : percent\_of\_income > 3: yes (14.3/5.6)  
## checking\_balance in {< 0 DM,> 200 DM,1 - 200 DM}:  
## :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (42.9/11.3)  
## savings\_balance = unknown:  
## :...credit\_history in {perfect,poor}: no (8.5)  
## : credit\_history in {critical,good,very good}:  
## : :...employment\_duration in {< 1 year,> 7 years,4 - 7 years,  
## : : unemployed}: no (52.3/17.3)  
## : employment\_duration = 1 - 4 years: yes (19.7/5.6)  
## savings\_balance = 100 - 500 DM:  
## :...existing\_loans\_count > 3: yes (3)  
## : existing\_loans\_count <= 3:  
## : :...credit\_history in {critical,poor,very good}: no (24.6/7.6)  
## : credit\_history = perfect: yes (2.4)  
## : credit\_history = good:  
## : :...months\_loan\_duration <= 27: no (23.7/10.5)  
## : months\_loan\_duration > 27: yes (5.6)  
## savings\_balance = < 100 DM:  
## :...months\_loan\_duration > 42: yes (28/5.2)  
## months\_loan\_duration <= 42:  
## :...percent\_of\_income <= 2:  
## :...employment\_duration in {1 - 4 years,4 - 7 years,  
## : : unemployed}: no (86.2/23.8)  
## : employment\_duration in {< 1 year,> 7 years}:  
## : :...housing = other: no (4.8/1.6)  
## : housing = rent: yes (10.7/2.4)  
## : housing = own:  
## : :...phone = yes: yes (12.9/4)  
## : phone = no:  
## : :...percent\_of\_income <= 1: no (7.1/0.8)  
## : percent\_of\_income > 1: yes (17.5/7.1)  
## percent\_of\_income > 2:  
## :...years\_at\_residence <= 1: no (31.6/8.5)  
## years\_at\_residence > 1:  
## :...credit\_history in {perfect,poor}: yes (20.9/1.6)  
## credit\_history in {critical,good,very good}:  
## :...job = skilled: yes (95/34.7)  
## job = unemployed: no (1.6)  
## job = management:  
## :...amount <= 11590: no (23.8/7)  
## : amount > 11590: yes (3.8)  
## job = unskilled:  
## :...checking\_balance in {< 0 DM,  
## : > 200 DM}: yes (23.8/9.5)  
## checking\_balance = 1 - 200 DM: no (17.9/6.2)  
##   
## ----- Trial 2: -----  
##   
## Decision tree:  
##   
## checking\_balance = unknown:  
## :...other\_credit = bank:  
## : :...existing\_loans\_count > 2: no (3.3)  
## : : existing\_loans\_count <= 2:  
## : : :...months\_loan\_duration <= 8: no (4)  
## : : months\_loan\_duration > 8: yes (43/16.6)  
## : other\_credit in {none,store}:  
## : :...employment\_duration in {< 1 year,unemployed}:  
## : :...purpose in {business,renovations}: yes (6.4)  
## : : purpose in {car,car0,education}: no (13.2)  
## : : purpose = furniture/appliances:  
## : : :...amount <= 4594: no (22.5/7.3)  
## : : amount > 4594: yes (9.1)  
## : employment\_duration in {> 7 years,1 - 4 years,4 - 7 years}:  
## : :...percent\_of\_income <= 3: no (92.7/3.6)  
## : percent\_of\_income > 3:  
## : :...age > 30: no (73.6/5.5)  
## : age <= 30:  
## : :...job in {management,unemployed,unskilled}: yes (14/4)  
## : job = skilled:  
## : :...credit\_history = very good: no (0)  
## : credit\_history = poor: yes (3.6)  
## : credit\_history in {critical,good,perfect}:  
## : :...age <= 29: no (20.4/4.6)  
## : age > 29: yes (2.7)  
## checking\_balance in {< 0 DM,> 200 DM,1 - 200 DM}:  
## :...housing = other:  
## :...dependents > 1: yes (28.3/7.6)  
## : dependents <= 1:  
## : :...employment\_duration in {< 1 year,4 - 7 years,  
## : : unemployed}: no (22.9/4.5)  
## : employment\_duration in {> 7 years,1 - 4 years}: yes (29.6/10.5)  
## housing = rent:  
## :...credit\_history = perfect: yes (5.3)  
## : credit\_history = poor: no (7.1/0.7)  
## : credit\_history in {critical,good,very good}:  
## : :...employment\_duration = < 1 year: yes (28.3/9.3)  
## : employment\_duration in {> 7 years,4 - 7 years,  
## : : unemployed}: no (33.9/12.3)  
## : employment\_duration = 1 - 4 years:  
## : :...checking\_balance = > 200 DM: no (2)  
## : checking\_balance in {< 0 DM,1 - 200 DM}:  
## : :...years\_at\_residence <= 3: no (10.3/3.8)  
## : years\_at\_residence > 3: yes (20.4/3.1)  
## housing = own:  
## :...job in {management,unemployed}: yes (55.8/19.8)  
## job in {skilled,unskilled}:  
## :...months\_loan\_duration <= 7: no (25.3/2)  
## months\_loan\_duration > 7:  
## :...years\_at\_residence > 3: no (92.2/29.6)  
## years\_at\_residence <= 3:  
## :...purpose = renovations: yes (7/1.3)  
## purpose in {business,car0,education}: no (32.2/5.3)  
## purpose = car:  
## :...months\_loan\_duration > 40: no (7.2/0.7)  
## : months\_loan\_duration <= 40:  
## : :...amount <= 947: yes (12.9)  
## : amount > 947:  
## : :...months\_loan\_duration <= 16: no (23.2/8.5)  
## : months\_loan\_duration > 16: [S1]  
## purpose = furniture/appliances:  
## :...savings\_balance in {> 1000 DM,unknown}: no (15.4/3.2)  
## savings\_balance in {100 - 500 DM,  
## : 500 - 1000 DM}: yes (14.6/4.5)  
## savings\_balance = < 100 DM:  
## :...months\_loan\_duration > 36: yes (7.1)  
## months\_loan\_duration <= 36:  
## :...existing\_loans\_count > 1: no (14.1/4.3)  
## existing\_loans\_count <= 1: [S2]  
##   
## SubTree [S1]  
##   
## savings\_balance in {< 100 DM,> 1000 DM,500 - 1000 DM,unknown}: yes (22.5/2.7)  
## savings\_balance = 100 - 500 DM: no (4.5/0.7)  
##   
## SubTree [S2]  
##   
## checking\_balance = < 0 DM: no (22.4/9.1)  
## checking\_balance in {> 200 DM,1 - 200 DM}: yes (46.7/20)  
##   
## ----- Trial 3: -----  
##   
## Decision tree:  
##   
## checking\_balance in {> 200 DM,unknown}:  
## :...employment\_duration = > 7 years: no (98.9/17.1)  
## : employment\_duration = unemployed: yes (16/6.7)  
## : employment\_duration = < 1 year:  
## : :...amount <= 1333: no (11.7)  
## : : amount > 1333:  
## : : :...amount <= 6681: no (38.2/16.3)  
## : : amount > 6681: yes (5.3)  
## : employment\_duration = 4 - 7 years:  
## : :...checking\_balance = > 200 DM: yes (9.6/3.6)  
## : : checking\_balance = unknown:  
## : : :...age <= 22: yes (6.5/1.6)  
## : : age > 22: no (42.6/1.5)  
## : employment\_duration = 1 - 4 years:  
## : :...percent\_of\_income <= 1: no (20.6/1.5)  
## : percent\_of\_income > 1:  
## : :...job in {skilled,unemployed}: no (64.9/17.6)  
## : job in {management,unskilled}:  
## : :...existing\_loans\_count > 2: yes (2.4)  
## : existing\_loans\_count <= 2:  
## : :...age <= 34: yes (26.4/10.7)  
## : age > 34: no (10.5)  
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (35.8/12)  
## savings\_balance = 100 - 500 DM:  
## :...amount <= 1285: yes (12.8/0.5)  
## : amount > 1285:  
## : :...existing\_loans\_count <= 1: no (27/9.2)  
## : existing\_loans\_count > 1: yes (15.8/4.9)  
## savings\_balance = unknown:  
## :...credit\_history in {critical,perfect,poor}: no (15.5)  
## : credit\_history in {good,very good}:  
## : :...age > 56: no (4.5)  
## : age <= 56:  
## : :...months\_loan\_duration <= 18: yes (24.5/5.6)  
## : months\_loan\_duration > 18: no (28.4/12.3)  
## savings\_balance = < 100 DM:  
## :...months\_loan\_duration <= 11:  
## :...job = management: yes (13.7/4.9)  
## : job in {skilled,unemployed,unskilled}: no (45.9/10)  
## months\_loan\_duration > 11:  
## :...percent\_of\_income <= 1:  
## :...credit\_history in {critical,poor,very good}: no (11.1)  
## : credit\_history in {good,perfect}: yes (24.4/11)  
## percent\_of\_income > 1:  
## :...job = unemployed: yes (7/3.1)  
## job = management:  
## :...years\_at\_residence <= 1: no (6.6)  
## : years\_at\_residence > 1:  
## : :...checking\_balance = < 0 DM: no (23.1/7)  
## : checking\_balance = 1 - 200 DM: yes (15.8/4)  
## job = unskilled:  
## :...housing in {other,rent}: yes (12.2/2.2)  
## : housing = own:  
## : :...purpose = car: yes (18.1/3.9)  
## : purpose in {business,car0,education,  
## : furniture/appliances,  
## : renovations}: no (32.1/11.1)  
## job = skilled:  
## :...checking\_balance = < 0 DM:  
## :...credit\_history in {poor,very good}: yes (16.6)  
## : credit\_history in {critical,good,perfect}:  
## : :...purpose in {business,car0,education,  
## : : renovations}: yes (10.2/1.5)  
## : purpose = car:  
## : :...age <= 51: yes (34.6/8.1)  
## : : age > 51: no (4.4)  
## : purpose = furniture/appliances:  
## : :...years\_at\_residence <= 1: no (4.4)  
## : years\_at\_residence > 1:  
## : :...other\_credit = bank: yes (2.4)  
## : other\_credit = store: no (0.5)  
## : other\_credit = none:  
## : :...amount <= 1743: no (11.5/2.4)  
## : amount > 1743: yes (29/6.6)  
## checking\_balance = 1 - 200 DM:  
## :...months\_loan\_duration > 36: yes (6.5)  
## months\_loan\_duration <= 36:  
## :...other\_credit in {bank,store}: yes (8/1.5)  
## other\_credit = none:  
## :...dependents > 1: yes (7.4/3.1)  
## dependents <= 1:  
## :...percent\_of\_income <= 2: no (12.7/1.1)  
## percent\_of\_income > 2: [S1]  
##   
## SubTree [S1]  
##   
## purpose in {business,renovations}: yes (3.9)  
## purpose in {car,car0,education,furniture/appliances}: no (19.8/6.1)  
##   
## ----- Trial 4: -----  
##   
## Decision tree:  
##   
## checking\_balance in {> 200 DM,unknown}:  
## :...other\_credit = store: no (20.6/9.6)  
## : other\_credit = none:  
## : :...employment\_duration in {> 7 years,1 - 4 years,4 - 7 years,  
## : : : unemployed}: no (211.3/45.7)  
## : : employment\_duration = < 1 year:  
## : : :...amount <= 1333: no (8.8)  
## : : amount > 1333:  
## : : :...purpose in {business,car0,education,furniture/appliances,  
## : : : renovations}: yes (32.9/8.1)  
## : : purpose = car: no (4.9)  
## : other\_credit = bank:  
## : :...age > 44: no (14.4/1.2)  
## : age <= 44:  
## : :...years\_at\_residence <= 1: no (5)  
## : years\_at\_residence > 1:  
## : :...housing = rent: yes (4.3)  
## : housing in {other,own}:  
## : :...job = unemployed: yes (0)  
## : job = management: no (4)  
## : job in {skilled,unskilled}:  
## : :...age <= 26: no (3.7)  
## : age > 26:  
## : :...savings\_balance in {< 100 DM,500 - 1000 DM,  
## : : unknown}: yes (30.6/7.4)  
## : savings\_balance in {> 1000 DM,  
## : 100 - 500 DM}: no (4)  
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...credit\_history = perfect:  
## :...housing in {other,rent}: yes (7.8)  
## : housing = own: no (20.5/9)  
## credit\_history = poor:  
## :...checking\_balance = < 0 DM: yes (10.4/2.2)  
## : checking\_balance = 1 - 200 DM:  
## : :...other\_credit in {bank,none}: no (24/4.3)  
## : other\_credit = store: yes (5.8/1.2)  
## credit\_history = very good:  
## :...age <= 23: no (5.7)  
## : age > 23:  
## : :...months\_loan\_duration <= 27: yes (28.4/3.7)  
## : months\_loan\_duration > 27: no (6.9/2)  
## credit\_history = critical:  
## :...years\_at\_residence <= 1: no (6.7)  
## : years\_at\_residence > 1:  
## : :...purpose in {business,car,car0,renovations}: no (62.2/21.9)  
## : purpose = education: yes (7.9/0.9)  
## : purpose = furniture/appliances:  
## : :...phone = yes: no (14.5/2.8)  
## : phone = no:  
## : :...amount <= 1175: no (5.2)  
## : amount > 1175: yes (30.1/7.6)  
## credit\_history = good:  
## :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (15.7/4.7)  
## savings\_balance = 100 - 500 DM: yes (32.1/11.7)  
## savings\_balance = unknown:  
## :...job = unskilled: no (4.4)  
## : job in {management,skilled,unemployed}:  
## : :...checking\_balance = < 0 DM: yes (27.8/6)  
## : checking\_balance = 1 - 200 DM: no (26.8/10.4)  
## savings\_balance = < 100 DM:  
## :...dependents > 1:  
## :...existing\_loans\_count > 1: no (2.6/0.4)  
## : existing\_loans\_count <= 1:  
## : :...years\_at\_residence <= 2: yes (10.2/2.9)  
## : years\_at\_residence > 2: no (20.4/5.9)  
## dependents <= 1:  
## :...purpose in {business,car0}: no (9.7/2.5)  
## purpose in {education,renovations}: yes (13/5.1)  
## purpose = car:  
## :...employment\_duration in {< 1 year,> 7 years,  
## : : 4 - 7 years}: yes (32/8.3)  
## : employment\_duration in {1 - 4 years,  
## : unemployed}: no (24.9/9)  
## purpose = furniture/appliances:  
## :...months\_loan\_duration > 39: yes (4.8)  
## months\_loan\_duration <= 39:  
## :...phone = yes: yes (21.9/9.2)  
## phone = no:  
## :...employment\_duration in {< 1 year,> 7 years,  
## : 4 - 7 years}: no (34.1/8.1)  
## employment\_duration = unemployed: yes (3.3/0.4)  
## employment\_duration = 1 - 4 years:  
## :...percent\_of\_income <= 1: yes (3.8)  
## percent\_of\_income > 1:  
## :...months\_loan\_duration > 21: no (4.9/0.4)  
## months\_loan\_duration <= 21:  
## :...years\_at\_residence <= 3: no (20.9/8.8)  
## years\_at\_residence > 3: yes (5.8)  
##   
## ----- Trial 5: -----  
##   
## Decision tree:  
##   
## checking\_balance = unknown:  
## :...other\_credit = store: yes (16.9/7.5)  
## : other\_credit = bank:  
## : :...housing = other: no (8.3/1.8)  
## : : housing = rent: yes (4.4/0.8)  
## : : housing = own:  
## : : :...phone = no: no (26.9/9.7)  
## : : phone = yes: yes (12.1/5)  
## : other\_credit = none:  
## : :...credit\_history in {critical,perfect,very good}: no (60.4/5.1)  
## : credit\_history in {good,poor}:  
## : :...purpose in {business,car,car0,education}: no (53.6/12.8)  
## : purpose = renovations: yes (7.3/1.1)  
## : purpose = furniture/appliances:  
## : :...job = unemployed: no (0)  
## : job in {management,unskilled}: yes (19.2/7)  
## : job = skilled:  
## : :...phone = yes: no (14.6/1.8)  
## : phone = no:  
## : :...age > 32: no (9.2)  
## : age <= 32:  
## : :...employment\_duration = 1 - 4 years: no (4.1)  
## : employment\_duration in {< 1 year,> 7 years,  
## : : 4 - 7 years,unemployed}:  
## : :...savings\_balance in {< 100 DM,  
## : : 100 - 500 DM}: yes (20.5/3)  
## : savings\_balance in {> 1000 DM,500 - 1000 DM,  
## : unknown}: no (3.4)  
## checking\_balance in {< 0 DM,> 200 DM,1 - 200 DM}:  
## :...percent\_of\_income <= 2:  
## :...amount > 11054: yes (14.2/1.2)  
## : amount <= 11054:  
## : :...other\_credit = bank: no (32.3/9.7)  
## : other\_credit = store: yes (8.9/2.6)  
## : other\_credit = none:  
## : :...purpose in {business,renovations}: yes (20.3/9.1)  
## : purpose in {car0,education}: no (8.4/3.7)  
## : purpose = car:  
## : :...savings\_balance in {< 100 DM,> 1000 DM,500 - 1000 DM,  
## : : : unknown}: no (46.6/7.9)  
## : : savings\_balance = 100 - 500 DM: yes (13.8/3.3)  
## : purpose = furniture/appliances:  
## : :...employment\_duration in {> 7 years,  
## : : 4 - 7 years}: no (18.2/2.6)  
## : employment\_duration in {1 - 4 years,  
## : : unemployed}: yes (50.8/19.5)  
## : employment\_duration = < 1 year:  
## : :...job in {management,skilled,unemployed}: no (16.3/2.9)  
## : job = unskilled: yes (6/1.6)  
## percent\_of\_income > 2:  
## :...years\_at\_residence <= 1:  
## :...other\_credit in {bank,store}: no (7.6)  
## : other\_credit = none:  
## : :...months\_loan\_duration > 42: no (2.9)  
## : months\_loan\_duration <= 42:  
## : :...age <= 36: no (26.6/8.4)  
## : age > 36: yes (5.3)  
## years\_at\_residence > 1:  
## :...job = unemployed: no (5.2)  
## job in {management,skilled,unskilled}:  
## :...credit\_history = perfect: yes (10.9)  
## credit\_history in {critical,good,poor,very good}:  
## :...employment\_duration = < 1 year:  
## :...checking\_balance = > 200 DM: no (2.7)  
## : checking\_balance in {< 0 DM,1 - 200 DM}:  
## : :...months\_loan\_duration > 21: yes (23.4/0.7)  
## : months\_loan\_duration <= 21:  
## : :...amount <= 1928: yes (18.4/4.4)  
## : amount > 1928: no (4.5)  
## employment\_duration in {> 7 years,1 - 4 years,4 - 7 years,  
## : unemployed}:  
## :...months\_loan\_duration <= 11:  
## :...age > 47: no (12.2)  
## : age <= 47:  
## : :...purpose in {business,car,car0,  
## : : furniture/appliances,  
## : : renovations}: no (25/9.2)  
## : purpose = education: yes (3.5)  
## months\_loan\_duration > 11:  
## :...savings\_balance in {> 1000 DM,100 - 500 DM}:  
## :...age <= 58: no (22.7/3.4)  
## : age > 58: yes (4.4)  
## savings\_balance in {< 100 DM,500 - 1000 DM,unknown}:  
## :...years\_at\_residence <= 2: yes (76.1/22.8)  
## years\_at\_residence > 2:  
## :...purpose in {business,car0,  
## : education}: yes (24.7/7.1)  
## purpose = renovations: no (1.1)  
## purpose = furniture/appliances: [S1]  
## purpose = car:  
## :...amount <= 1388: yes (17.8/2.2)  
## amount > 1388:  
## :...housing = own: no (10.9)  
## housing in {other,rent}: [S2]  
##   
## SubTree [S1]  
##   
## employment\_duration = unemployed: no (4.4)  
## employment\_duration in {> 7 years,1 - 4 years,4 - 7 years}:  
## :...checking\_balance = < 0 DM: yes (35.6/12.4)  
## checking\_balance in {> 200 DM,1 - 200 DM}: no (29/10.5)  
##   
## SubTree [S2]  
##   
## savings\_balance in {< 100 DM,500 - 1000 DM}: yes (21.4/6.4)  
## savings\_balance = unknown: no (6.8/1.5)  
##   
## ----- Trial 6: -----  
##   
## Decision tree:  
##   
## checking\_balance in {> 200 DM,unknown}:  
## :...purpose = car0: no (2.2)  
## : purpose = renovations: yes (8.4/3.3)  
## : purpose = education:  
## : :...age <= 44: yes (19.8/7.7)  
## : : age > 44: no (4.4)  
## : purpose = business:  
## : :...existing\_loans\_count > 2: yes (3.3)  
## : : existing\_loans\_count <= 2:  
## : : :...amount <= 1823: no (8.1)  
## : : amount > 1823:  
## : : :...percent\_of\_income <= 3: no (12.1/3.3)  
## : : percent\_of\_income > 3: yes (13.2/3.4)  
## : purpose = car:  
## : :...job in {management,unemployed}: no (20.8/1.6)  
## : : job = unskilled:  
## : : :...years\_at\_residence <= 3: no (11/1.3)  
## : : : years\_at\_residence > 3: yes (14.5/3.2)  
## : : job = skilled:  
## : : :...other\_credit in {bank,store}: yes (17.6/4.9)  
## : : other\_credit = none:  
## : : :...existing\_loans\_count <= 2: no (24.6)  
## : : existing\_loans\_count > 2: yes (2.4/0.3)  
## : purpose = furniture/appliances:  
## : :...age > 44: no (22.7)  
## : age <= 44:  
## : :...job = unemployed: no (0)  
## : job = unskilled:  
## : :...existing\_loans\_count <= 1: yes (20.9/5.6)  
## : : existing\_loans\_count > 1: no (4.5)  
## : job in {management,skilled}:  
## : :...dependents > 1: no (6.6)  
## : dependents <= 1:  
## : :...existing\_loans\_count <= 1:  
## : :...savings\_balance in {> 1000 DM,100 - 500 DM,  
## : : : 500 - 1000 DM,  
## : : : unknown}: no (16.9)  
## : : savings\_balance = < 100 DM:  
## : : :...age <= 22: yes (8.5/1.3)  
## : : age > 22: no (43.1/8.8)  
## : existing\_loans\_count > 1:  
## : :...housing in {other,rent}: yes (9.9/2.1)  
## : housing = own:  
## : :...credit\_history in {critical,poor,  
## : : very good}: no (18.6/1.6)  
## : credit\_history in {good,perfect}: yes (14.9/4.3)  
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...credit\_history = perfect: yes (28.1/9.6)  
## credit\_history = very good:  
## :...age <= 23: no (5.5)  
## : age > 23: yes (30/8.1)  
## credit\_history = poor:  
## :...percent\_of\_income <= 1: no (6.5)  
## : percent\_of\_income > 1:  
## : :...savings\_balance in {500 - 1000 DM,unknown}: no (6.4)  
## : savings\_balance in {< 100 DM,> 1000 DM,100 - 500 DM}:  
## : :...dependents <= 1: yes (25.1/8)  
## : dependents > 1: no (5/0.9)  
## credit\_history = critical:  
## :...savings\_balance = unknown: no (8.4)  
## : savings\_balance in {< 100 DM,> 1000 DM,100 - 500 DM,500 - 1000 DM}:  
## : :...other\_credit = bank: yes (16.2/4.3)  
## : other\_credit = store: no (3.7/0.9)  
## : other\_credit = none:  
## : :...savings\_balance in {> 1000 DM,500 - 1000 DM}: yes (7.3/2.3)  
## : savings\_balance = 100 - 500 DM: no (5.9)  
## : savings\_balance = < 100 DM:  
## : :...purpose = business: no (4.5/2.2)  
## : purpose in {car0,education,renovations}: yes (8.5/2.2)  
## : purpose = car:  
## : :...age <= 29: yes (6.9)  
## : : age > 29: no (25.6/6.9)  
## : purpose = furniture/appliances:  
## : :...months\_loan\_duration <= 36: no (38.4/10.9)  
## : months\_loan\_duration > 36: yes (3.8)  
## credit\_history = good:  
## :...amount > 8086: yes (24/3.8)  
## amount <= 8086:  
## :...phone = yes:  
## :...age <= 28: yes (23.9/7.5)  
## : age > 28: no (69.4/17.9)  
## phone = no:  
## :...other\_credit in {bank,store}: yes (25.1/7.2)  
## other\_credit = none:  
## :...percent\_of\_income <= 2:  
## :...job in {management,unemployed,unskilled}: no (15.6/2.7)  
## : job = skilled:  
## : :...amount <= 1386: yes (9.9/1)  
## : amount > 1386:  
## : :...age <= 24: yes (13.4/4.6)  
## : age > 24: no (27.8/3.1)  
## percent\_of\_income > 2:  
## :...checking\_balance = < 0 DM: yes (62.5/21.4)  
## checking\_balance = 1 - 200 DM:  
## :...months\_loan\_duration > 42: yes (4.9)  
## months\_loan\_duration <= 42:  
## :...existing\_loans\_count > 1: no (5)  
## existing\_loans\_count <= 1:  
## :...age <= 35: no (39.4/13.2)  
## age > 35: yes (14.7/4.2)  
##   
## ----- Trial 7: -----  
##   
## Decision tree:  
##   
## checking\_balance = unknown:  
## :...employment\_duration in {> 7 years,4 - 7 years}: no (101.1/20.4)  
## : employment\_duration = unemployed: yes (16.6/8)  
## : employment\_duration = < 1 year:  
## : :...amount <= 4594: no (30/5.7)  
## : : amount > 4594: yes (10.6/0.3)  
## : employment\_duration = 1 - 4 years:  
## : :...dependents > 1: no (8)  
## : dependents <= 1:  
## : :...months\_loan\_duration <= 16: no (32.8/5.3)  
## : months\_loan\_duration > 16:  
## : :...existing\_loans\_count > 2: yes (2.7)  
## : existing\_loans\_count <= 2:  
## : :...percent\_of\_income <= 3: no (20.9/5.9)  
## : percent\_of\_income > 3:  
## : :...purpose in {business,car0,education}: yes (10.8)  
## : purpose in {car,furniture/appliances,  
## : renovations}: no (19.7/7.5)  
## checking\_balance in {< 0 DM,> 200 DM,1 - 200 DM}:  
## :...purpose in {car0,education,renovations}: no (67.2/29.2)  
## purpose = business:  
## :...age > 46: yes (5.2)  
## : age <= 46:  
## : :...months\_loan\_duration <= 18: no (17.5)  
## : months\_loan\_duration > 18:  
## : :...other\_credit in {bank,store}: no (10/0.5)  
## : other\_credit = none:  
## : :...employment\_duration in {> 7 years,  
## : : unemployed}: yes (6.6)  
## : employment\_duration in {< 1 year,1 - 4 years,4 - 7 years}:  
## : :...age <= 25: yes (4)  
## : age > 25: no (19.2/5.6)  
## purpose = car:  
## :...amount <= 1297: yes (52.4/12.9)  
## : amount > 1297:  
## : :...percent\_of\_income <= 2:  
## : :...phone = no: no (32.7/6.1)  
## : : phone = yes:  
## : : :...years\_at\_residence <= 3: no (20/4.9)  
## : : years\_at\_residence > 3: yes (14.7/3.8)  
## : percent\_of\_income > 2:  
## : :...percent\_of\_income <= 3: yes (33.1/11.3)  
## : percent\_of\_income > 3:  
## : :...months\_loan\_duration <= 18: no (18.2/1.6)  
## : months\_loan\_duration > 18:  
## : :...existing\_loans\_count <= 1: no (19.5/7.2)  
## : existing\_loans\_count > 1: yes (13.8/1)  
## purpose = furniture/appliances:  
## :...savings\_balance = > 1000 DM: no (5.2)  
## savings\_balance = 100 - 500 DM: yes (18.6/6)  
## savings\_balance in {< 100 DM,500 - 1000 DM,unknown}:  
## :...existing\_loans\_count > 1:  
## :...existing\_loans\_count > 2: no (3.6)  
## : existing\_loans\_count <= 2:  
## : :...housing = other: yes (3.3)  
## : housing in {own,rent}:  
## : :...savings\_balance = 500 - 1000 DM: yes (3.5/1)  
## : savings\_balance = unknown: no (6.9)  
## : savings\_balance = < 100 DM:  
## : :...age > 54: yes (2.1)  
## : age <= 54: [S1]  
## existing\_loans\_count <= 1:  
## :...credit\_history in {critical,perfect}: yes (20.3/7.6)  
## credit\_history in {poor,very good}: no (20.8/9.5)  
## credit\_history = good:  
## :...months\_loan\_duration <= 7: no (11.4)  
## months\_loan\_duration > 7:  
## :...other\_credit = bank: no (14.2/4.6)  
## other\_credit = store: yes (11.7/3.9)  
## other\_credit = none:  
## :...percent\_of\_income <= 1: no (20.5/5.2)  
## percent\_of\_income > 1:  
## :...amount > 6078: yes (10.9/1.1)  
## amount <= 6078:  
## :...dependents > 1: yes (8.7/2.5)  
## dependents <= 1: [S2]  
##   
## SubTree [S1]  
##   
## employment\_duration in {< 1 year,4 - 7 years}: yes (15/2.5)  
## employment\_duration in {> 7 years,1 - 4 years,unemployed}: no (25.7/2.9)  
##   
## SubTree [S2]  
##   
## employment\_duration = > 7 years: no (17.9/2.5)  
## employment\_duration in {< 1 year,1 - 4 years,4 - 7 years,unemployed}:  
## :...job = management: no (6.6)  
## job = unemployed: yes (1.1)  
## job in {skilled,unskilled}:  
## :...years\_at\_residence <= 1: no (11.8/1.8)  
## years\_at\_residence > 1:  
## :...checking\_balance = > 200 DM: no (14.7/6.3)  
## checking\_balance = 1 - 200 DM: yes (25.1/8.8)  
## checking\_balance = < 0 DM:  
## :...months\_loan\_duration <= 16: no (13.8/3.4)  
## months\_loan\_duration > 16: yes (19.1/5.5)  
##   
## ----- Trial 8: -----  
##   
## Decision tree:  
##   
## checking\_balance in {< 0 DM,1 - 200 DM}:  
## :...credit\_history = perfect:  
## : :...housing in {other,rent}: yes (8.3)  
## : : housing = own:  
## : : :...age <= 34: no (16.6/4.7)  
## : : age > 34: yes (5.8)  
## : credit\_history = poor:  
## : :...checking\_balance = < 0 DM: yes (12/2.7)  
## : : checking\_balance = 1 - 200 DM:  
## : : :...housing = rent: no (8.6)  
## : : housing in {other,own}:  
## : : :...amount <= 2279: yes (6.8/0.6)  
## : : amount > 2279: no (20/5.7)  
## : credit\_history = very good:  
## : :...existing\_loans\_count > 1: yes (2.5)  
## : : existing\_loans\_count <= 1:  
## : : :...age <= 23: no (3.7)  
## : : age > 23:  
## : : :...amount <= 8386: yes (32.9/8.1)  
## : : amount > 8386: no (2.5)  
## : credit\_history = critical:  
## : :...years\_at\_residence <= 1: no (8)  
## : : years\_at\_residence > 1:  
## : : :...savings\_balance in {> 1000 DM,100 - 500 DM,500 - 1000 DM,  
## : : : unknown}: no (25.5/5.7)  
## : : savings\_balance = < 100 DM:  
## : : :...age > 61: no (6)  
## : : age <= 61:  
## : : :...existing\_loans\_count > 2: no (10.7/2.4)  
## : : existing\_loans\_count <= 2:  
## : : :...age > 56: yes (5.4)  
## : : age <= 56:  
## : : :...amount > 2483: yes (34.1/8.9)  
## : : amount <= 2483:  
## : : :...purpose in {business,education}: yes (4.4)  
## : : purpose in {car,car0,furniture/appliances,  
## : : renovations}: no (41.4/10.8)  
## : credit\_history = good:  
## : :...amount > 8086: yes (26.6/4.8)  
## : amount <= 8086:  
## : :...savings\_balance in {> 1000 DM,500 - 1000 DM}: no (17.5/5.1)  
## : savings\_balance = 100 - 500 DM:  
## : :...months\_loan\_duration <= 27: no (21.3/7.1)  
## : : months\_loan\_duration > 27: yes (5.1)  
## : savings\_balance = unknown:  
## : :...age <= 56: yes (44.7/16.9)  
## : : age > 56: no (4.4)  
## : savings\_balance = < 100 DM:  
## : :...job = unemployed: yes (0.9)  
## : job = management:  
## : :...employment\_duration in {< 1 year,1 - 4 years,4 - 7 years,  
## : : : unemployed}: no (17.3/1.6)  
## : : employment\_duration = > 7 years: yes (8/1.2)  
## : job = unskilled:  
## : :...months\_loan\_duration <= 26: no (59/19.7)  
## : : months\_loan\_duration > 26: yes (3.3)  
## : job = skilled:  
## : :...purpose in {business,car0,education,  
## : : renovations}: yes (16.6/4.1)  
## : purpose = car:  
## : :...dependents <= 1: yes (27.7/10.6)  
## : : dependents > 1: no (8.1/1.4)  
## : purpose = furniture/appliances:  
## : :...years\_at\_residence <= 1: no (18.7/6.5)  
## : years\_at\_residence > 1:  
## : :...other\_credit = bank: yes (4.5)  
## : other\_credit = store: no (2.3)  
## : other\_credit = none:  
## : :...percent\_of\_income <= 3: yes (33.5/15)  
## : percent\_of\_income > 3: no (27.3/9.3)  
## checking\_balance in {> 200 DM,unknown}:  
## :...years\_at\_residence > 2: no (135.6/32.2)  
## years\_at\_residence <= 2:  
## :...months\_loan\_duration <= 8: no (12.9)  
## months\_loan\_duration > 8:  
## :...months\_loan\_duration <= 9: yes (10.4/1.3)  
## months\_loan\_duration > 9:  
## :...months\_loan\_duration <= 16: no (31.3/4.2)  
## months\_loan\_duration > 16:  
## :...purpose in {business,car0,renovations}: no (21.3/8.4)  
## purpose = education: yes (6.3/0.8)  
## purpose = car:  
## :...credit\_history in {critical,very good}: yes (17.3/2.6)  
## : credit\_history in {good,perfect,poor}: no (9.6)  
## purpose = furniture/appliances:  
## :...credit\_history in {critical,perfect,  
## : very good}: no (5.6)  
## credit\_history = poor: yes (4.9)  
## credit\_history = good:  
## :...housing in {other,rent}: no (2.6)  
## housing = own:  
## :...age <= 25: no (6.8)  
## age > 25: yes (29.2/10.2)  
##   
## ----- Trial 9: -----  
##   
## Decision tree:  
##   
## checking\_balance = unknown:  
## :...dependents > 1: no (26)  
## : dependents <= 1:  
## : :...amount <= 1474: no (39.7)  
## : amount > 1474:  
## : :...employment\_duration in {> 7 years,4 - 7 years}:  
## : :...years\_at\_residence > 2: no (21.8)  
## : : years\_at\_residence <= 2:  
## : : :...age <= 23: yes (4.1)  
## : : age > 23: no (19.7/4.2)  
## : employment\_duration in {< 1 year,1 - 4 years,unemployed}:  
## : :...purpose in {business,renovations}: yes (23.2/3.6)  
## : purpose in {car,car0,education,furniture/appliances}:  
## : :...other\_credit in {bank,store}: yes (29.1/10.5)  
## : other\_credit = none:  
## : :...purpose in {car,car0}: no (12.3)  
## : purpose in {education,furniture/appliances}:  
## : :...amount <= 4455: no (23.7/4.4)  
## : amount > 4455: yes (11.1/1.3)  
## checking\_balance in {< 0 DM,> 200 DM,1 - 200 DM}:  
## :...percent\_of\_income <= 2:  
## :...amount > 11054: yes (15.7/3.6)  
## : amount <= 11054:  
## : :...savings\_balance in {> 1000 DM,500 - 1000 DM,  
## : : unknown}: no (41.5/11.2)  
## : savings\_balance = 100 - 500 DM:  
## : :...other\_credit = bank: no (5.1)  
## : : other\_credit in {none,store}: yes (21.7/9.4)  
## : savings\_balance = < 100 DM:  
## : :...employment\_duration in {> 7 years,unemployed}: no (34.6/11.5)  
## : employment\_duration = 1 - 4 years:  
## : :...job = management: yes (5.1/0.8)  
## : : job in {skilled,unemployed,unskilled}: no (65.4/15.8)  
## : employment\_duration = < 1 year:  
## : :...amount <= 2327:  
## : : :...age <= 34: yes (20.5/1.9)  
## : : : age > 34: no (3)  
## : : amount > 2327:  
## : : :...other\_credit = bank: yes (2.8)  
## : : other\_credit in {none,store}: no (20.1/3.9)  
## : employment\_duration = 4 - 7 years:  
## : :...dependents > 1: no (4.6)  
## : dependents <= 1:  
## : :...amount <= 6527: no (16.8/7.2)  
## : amount > 6527: yes (7)  
## percent\_of\_income > 2:  
## :...housing = rent:  
## :...checking\_balance in {< 0 DM,1 - 200 DM}: yes (69/22.1)  
## : checking\_balance = > 200 DM: no (3.4)  
## housing = other:  
## :...existing\_loans\_count > 1: yes (18.7/5.3)  
## : existing\_loans\_count <= 1:  
## : :...savings\_balance in {< 100 DM,> 1000 DM,  
## : : 500 - 1000 DM}: yes (29.1/8.6)  
## : savings\_balance in {100 - 500 DM,unknown}: no (15.3/3.2)  
## housing = own:  
## :...credit\_history in {perfect,poor}: yes (26.9/7.4)  
## credit\_history = very good: no (14.9/5.6)  
## credit\_history = critical:  
## :...other\_credit = bank: yes (11.7/3.4)  
## : other\_credit in {none,store}: no (63/20.3)  
## credit\_history = good:  
## :...other\_credit = store: yes (8.9/1.4)  
## other\_credit in {bank,none}:  
## :...age > 54: no (9.5)  
## age <= 54:  
## :...existing\_loans\_count > 1: no (10.2/2.7)  
## existing\_loans\_count <= 1:  
## :...purpose in {business,renovations}: no (10.1/3.6)  
## purpose in {car0,education}: yes (4.7)  
## purpose = car:  
## :...other\_credit = bank: yes (4.9)  
## : other\_credit = none:  
## : :...years\_at\_residence > 2: no (14.8/4.5)  
## : years\_at\_residence <= 2:  
## : :...amount <= 2150: no (14.9/6.2)  
## : amount > 2150: yes (11.1)  
## purpose = furniture/appliances:  
## :...savings\_balance = 100 - 500 DM: yes (3.8)  
## savings\_balance in {> 1000 DM,  
## : 500 - 1000 DM}: no (2.8)  
## savings\_balance in {< 100 DM,unknown}:  
## :...months\_loan\_duration > 39: yes (3.3)  
## months\_loan\_duration <= 39:  
## :...dependents <= 1: no (57.6/19.4)  
## dependents > 1: yes (4.6/1.1)  
##   
##   
## Evaluation on training data (900 cases):  
##   
## Trial Decision Tree   
## ----- ----------------   
## Size Errors   
##   
## 0 56 133(14.8%)  
## 1 34 211(23.4%)  
## 2 39 201(22.3%)  
## 3 47 179(19.9%)  
## 4 46 174(19.3%)  
## 5 50 197(21.9%)  
## 6 55 187(20.8%)  
## 7 50 190(21.1%)  
## 8 51 192(21.3%)  
## 9 47 169(18.8%)  
## boost 34( 3.8%) <<  
##   
##   
## (a) (b) <-classified as  
## ---- ----  
## 629 4 (a): class no  
## 30 237 (b): class yes  
##   
##   
## Attribute usage:  
##   
## 100.00% checking\_balance  
## 100.00% purpose  
## 97.11% years\_at\_residence  
## 96.67% employment\_duration  
## 94.78% credit\_history  
## 94.67% other\_credit  
## 92.56% job  
## 92.11% percent\_of\_income  
## 90.33% amount  
## 85.11% months\_loan\_duration  
## 82.78% age  
## 82.78% existing\_loans\_count  
## 75.78% dependents  
## 71.56% housing  
## 70.78% savings\_balance  
## 49.22% phone  
##   
##   
## Time: 0.1 secs

credit\_boost\_pred10 <- predict(credit\_boost10, credit\_test)  
CrossTable(credit\_test$default, credit\_boost\_pred10,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 100   
##   
##   
## | predicted default   
## actual default | no | yes | Row Total |   
## ---------------|-----------|-----------|-----------|  
## no | 62 | 5 | 67 |   
## | 0.620 | 0.050 | |   
## ---------------|-----------|-----------|-----------|  
## yes | 13 | 20 | 33 |   
## | 0.130 | 0.200 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 75 | 25 | 100 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Accuracy of decision tree with using boosted trail is 82%, which is greater than previous.

Making mistakes more costlier than others. reduce the number of false negatives may be to reject a larger number of borderline applicants.

#Creating dimensions for cost matrix  
matrix\_dimensions <- list(c("no", "yes"), c("no", "yes"))  
names(matrix\_dimensions) <- c("predicted", "actual")  
matrix\_dimensions

## $predicted  
## [1] "no" "yes"  
##   
## $actual  
## [1] "no" "yes"

# build the matrix  
error\_cost <- matrix(c(0, 1, 4, 0), nrow = 2, dimnames = matrix\_dimensions)  
error\_cost

## actual  
## predicted no yes  
## no 0 4  
## yes 1 0

# apply the cost matrix to the tree  
credit\_cost <- C5.0(credit\_train[-17], credit\_train$default,  
 costs = error\_cost)  
credit\_cost\_pred <- predict(credit\_cost, credit\_test)  
  
CrossTable(credit\_test$default, credit\_cost\_pred,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 100   
##   
##   
## | predicted default   
## actual default | no | yes | Row Total |   
## ---------------|-----------|-----------|-----------|  
## no | 37 | 30 | 67 |   
## | 0.370 | 0.300 | |   
## ---------------|-----------|-----------|-----------|  
## yes | 7 | 26 | 33 |   
## | 0.070 | 0.260 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 44 | 56 | 100 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Compared to our boosted model, this version makes more mistakes overall: 37 percent error here versus 18 percent in the boosted case. However, the types of mistakes are very different. This trade resulting in a reduction of false negatives at the expense of increasing false positives may be acceptable if our cost estimates were accurate.

**Rule Learners...........**

**Step 1: Collecting the data**

source :UCI Machine Learning Repository. The dataset includes information on 8,124 mushroom samples from 23 species of gilled mushrooms listed in Audubon Society Field Guide to North American Mushrooms (1981). The 22 features of the mushroom samples, including characteristics such as cap shape, cap color, odor, gill size and color, stalk shape, and habitat.

**Step 2: Exploring and preparing the data**

mushrooms <- read.csv("mushrooms.csv", stringsAsFactors = TRUE)

#Examine the structure of the data frame   
str(mushrooms)

## 'data.frame': 8124 obs. of 23 variables:  
## $ type : Factor w/ 2 levels "edible","poisonous": 2 1 1 2 1 1 1 1 2 1 ...  
## $ cap\_shape : Factor w/ 6 levels "bell","conical",..: 3 3 1 3 3 3 1 1 3 1 ...  
## $ cap\_surface : Factor w/ 4 levels "fibrous","grooves",..: 4 4 4 3 4 3 4 3 3 4 ...  
## $ cap\_color : Factor w/ 10 levels "brown","buff",..: 1 10 9 9 4 10 9 9 9 10 ...  
## $ bruises : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2 2 ...  
## $ odor : Factor w/ 9 levels "almond","anise",..: 8 1 2 8 7 1 1 2 8 1 ...  
## $ gill\_attachment : Factor w/ 2 levels "attached","free": 2 2 2 2 2 2 2 2 2 2 ...  
## $ gill\_spacing : Factor w/ 2 levels "close","crowded": 1 1 1 1 2 1 1 1 1 1 ...  
## $ gill\_size : Factor w/ 2 levels "broad","narrow": 2 1 1 2 1 1 1 1 2 1 ...  
## $ gill\_color : Factor w/ 12 levels "black","brown",..: 1 1 2 2 1 2 5 2 8 5 ...  
## $ stalk\_shape : Factor w/ 2 levels "enlarging","tapering": 1 1 1 1 2 1 1 1 1 1 ...  
## $ stalk\_root : Factor w/ 5 levels "bulbous","club",..: 3 2 2 3 3 2 2 2 3 2 ...  
## $ stalk\_surface\_above\_ring: Factor w/ 4 levels "fibrous","scaly",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ stalk\_surface\_below\_ring: Factor w/ 4 levels "fibrous","scaly",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ stalk\_color\_above\_ring : Factor w/ 9 levels "brown","buff",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ stalk\_color\_below\_ring : Factor w/ 9 levels "brown","buff",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ veil\_type : Factor w/ 1 level "partial": 1 1 1 1 1 1 1 1 1 1 ...  
## $ veil\_color : Factor w/ 4 levels "brown","orange",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ ring\_number : Factor w/ 3 levels "none","one","two": 2 2 2 2 2 2 2 2 2 2 ...  
## $ ring\_type : Factor w/ 5 levels "evanescent","flaring",..: 5 5 5 5 1 5 5 5 5 5 ...  
## $ spore\_print\_color : Factor w/ 9 levels "black","brown",..: 1 2 2 1 2 1 1 2 1 1 ...  
## $ population : Factor w/ 6 levels "abundant","clustered",..: 4 3 3 4 1 3 3 4 5 4 ...  
## $ habitat : Factor w/ 7 levels "grasses","leaves",..: 5 1 3 5 1 1 3 3 1 3 ...

veil\_type variable has only one level. In any case, since the veil type does not vary across samples, it does not provide any useful information for prediction. We will drop this variable from our analysis using the following command:

#Droping veil type variable   
mushrooms$veil\_type <- NULL

#To be sure veil type is deleted  
str(mushrooms)

## 'data.frame': 8124 obs. of 22 variables:  
## $ type : Factor w/ 2 levels "edible","poisonous": 2 1 1 2 1 1 1 1 2 1 ...  
## $ cap\_shape : Factor w/ 6 levels "bell","conical",..: 3 3 1 3 3 3 1 1 3 1 ...  
## $ cap\_surface : Factor w/ 4 levels "fibrous","grooves",..: 4 4 4 3 4 3 4 3 3 4 ...  
## $ cap\_color : Factor w/ 10 levels "brown","buff",..: 1 10 9 9 4 10 9 9 9 10 ...  
## $ bruises : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2 2 ...  
## $ odor : Factor w/ 9 levels "almond","anise",..: 8 1 2 8 7 1 1 2 8 1 ...  
## $ gill\_attachment : Factor w/ 2 levels "attached","free": 2 2 2 2 2 2 2 2 2 2 ...  
## $ gill\_spacing : Factor w/ 2 levels "close","crowded": 1 1 1 1 2 1 1 1 1 1 ...  
## $ gill\_size : Factor w/ 2 levels "broad","narrow": 2 1 1 2 1 1 1 1 2 1 ...  
## $ gill\_color : Factor w/ 12 levels "black","brown",..: 1 1 2 2 1 2 5 2 8 5 ...  
## $ stalk\_shape : Factor w/ 2 levels "enlarging","tapering": 1 1 1 1 2 1 1 1 1 1 ...  
## $ stalk\_root : Factor w/ 5 levels "bulbous","club",..: 3 2 2 3 3 2 2 2 3 2 ...  
## $ stalk\_surface\_above\_ring: Factor w/ 4 levels "fibrous","scaly",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ stalk\_surface\_below\_ring: Factor w/ 4 levels "fibrous","scaly",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ stalk\_color\_above\_ring : Factor w/ 9 levels "brown","buff",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ stalk\_color\_below\_ring : Factor w/ 9 levels "brown","buff",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ veil\_color : Factor w/ 4 levels "brown","orange",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ ring\_number : Factor w/ 3 levels "none","one","two": 2 2 2 2 2 2 2 2 2 2 ...  
## $ ring\_type : Factor w/ 5 levels "evanescent","flaring",..: 5 5 5 5 1 5 5 5 5 5 ...  
## $ spore\_print\_color : Factor w/ 9 levels "black","brown",..: 1 2 2 1 2 1 1 2 1 1 ...  
## $ population : Factor w/ 6 levels "abundant","clustered",..: 4 3 3 4 1 3 3 4 5 4 ...  
## $ habitat : Factor w/ 7 levels "grasses","leaves",..: 5 1 3 5 1 1 3 3 1 3 ...

veil type has been deleted.

# Examine the type of mushroom  
table(mushrooms$type)

##   
## edible poisonous   
## 4208 3916

#Creating random sample   
set.seed(123)  
train\_sample <- sample(8124, 7000)

#Sructure of random sample  
str(train\_sample)

## int [1:7000] 2337 6404 3322 7171 7637 370 4288 7244 4476 3706 ...

# split the data frames  
mushrooms\_train <- mushrooms[train\_sample, ]  
mushrooms\_test <- mushrooms[-train\_sample, ]

**Step 3: Training a model on the data**

library(RWeka)

#using 1R algorithm  
mushroom\_1R <- OneR(type ~ ., data = mushrooms)

**Step 4: Evaluating model performance**

mushroom\_1R

## odor:  
## almond -> edible  
## anise -> edible  
## creosote -> poisonous  
## fishy -> poisonous  
## foul -> poisonous  
## musty -> poisonous  
## none -> edible  
## pungent -> poisonous  
## spicy -> poisonous  
## (8004/8124 instances correct)

The odor feature was selected for rule generation.The categories of odor, such as almond, anise, and so on, specify rules for whether the mushroom is likely to be edible or poisonous.

summary(mushroom\_1R)

##   
## === Summary ===  
##   
## Correctly Classified Instances 8004 98.5229 %  
## Incorrectly Classified Instances 120 1.4771 %  
## Kappa statistic 0.9704  
## Mean absolute error 0.0148  
## Root mean squared error 0.1215  
## Relative absolute error 2.958 %  
## Root relative squared error 24.323 %  
## Total Number of Instances 8124   
##   
## === Confusion Matrix ===  
##   
## a b <-- classified as  
## 4208 0 | a = edible  
## 120 3796 | b = poisonous

It classified 120 poisonous mushroom as edible, which is very dengerous.

mushroom\_pred <- predict(mushroom\_1R, mushrooms\_test)

# cross tabulation of predicted versus actual classes  
library(gmodels)  
CrossTable(mushrooms\_test$type, mushroom\_pred,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1124   
##   
##   
## | predicted default   
## actual default | edible | poisonous | Row Total |   
## ---------------|-----------|-----------|-----------|  
## edible | 582 | 0 | 582 |   
## | 0.518 | 0.000 | |   
## ---------------|-----------|-----------|-----------|  
## poisonous | 15 | 527 | 542 |   
## | 0.013 | 0.469 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 597 | 527 | 1124 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Accuracy is 98.66%.

**Step 5: Improving model performance**

I will use JRip(), a Java-based implementation of the RIPPER rule learning algorithm.

mushroom\_JRip <- JRip(type ~ ., data = mushrooms)

mushroom\_JRip

## JRIP rules:  
## ===========  
##   
## (odor = foul) => type=poisonous (2160.0/0.0)  
## (gill\_size = narrow) and (gill\_color = buff) => type=poisonous (1152.0/0.0)  
## (gill\_size = narrow) and (odor = pungent) => type=poisonous (256.0/0.0)  
## (odor = creosote) => type=poisonous (192.0/0.0)  
## (spore\_print\_color = green) => type=poisonous (72.0/0.0)  
## (stalk\_surface\_below\_ring = scaly) and (stalk\_surface\_above\_ring = silky) => type=poisonous (68.0/0.0)  
## (habitat = leaves) and (cap\_color = white) => type=poisonous (8.0/0.0)  
## (stalk\_color\_above\_ring = yellow) => type=poisonous (8.0/0.0)  
## => type=edible (4208.0/0.0)  
##   
## Number of Rules : 9

. If the odor is foul, then the mushroom type is poisonous . If the gill size is narrow and the gill color is buff, then the mushroom type is poisonous . If the gill size is narrow and the odor is pungent, then the mushroom type is poisonous so on.

summary(mushroom\_JRip)

##   
## === Summary ===  
##   
## Correctly Classified Instances 8124 100 %  
## Incorrectly Classified Instances 0 0 %  
## Kappa statistic 1   
## Mean absolute error 0   
## Root mean squared error 0   
## Relative absolute error 0 %  
## Root relative squared error 0 %  
## Total Number of Instances 8124   
##   
## === Confusion Matrix ===  
##   
## a b <-- classified as  
## 4208 0 | a = edible  
## 0 3916 | b = poisonous

mushroom\_pred <- predict(mushroom\_JRip, mushrooms\_test)

# cross tabulation of predicted versus actual classes  
library(gmodels)  
CrossTable(mushrooms\_test$type, mushroom\_pred,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1124   
##   
##   
## | predicted default   
## actual default | edible | poisonous | Row Total |   
## ---------------|-----------|-----------|-----------|  
## edible | 582 | 0 | 582 |   
## | 0.518 | 0.000 | |   
## ---------------|-----------|-----------|-----------|  
## poisonous | 0 | 542 | 542 |   
## | 0.000 | 0.482 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 582 | 542 | 1124 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Acuuracy is 100%.

# Rule Learner Using C5.0 Decision Trees (not in text)  
library(C50)  
mushroom\_c5rules <- C5.0(type ~ odor + gill\_size, data = mushrooms, rules = TRUE)  
summary(mushroom\_c5rules)

##   
## Call:  
## C5.0.formula(formula = type ~ odor + gill\_size, data = mushrooms, rules  
## = TRUE)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Mon May 01 07:39:55 2017  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 8124 cases (3 attributes) from undefined.data  
##   
## Rules:  
##   
## Rule 1: (4328/120, lift 1.9)  
## odor in {almond, anise, none}  
## -> class edible [0.972]  
##   
## Rule 2: (3796, lift 2.1)  
## odor in {creosote, fishy, foul, musty, pungent, spicy}  
## -> class poisonous [1.000]  
##   
## Default class: edible  
##   
##   
## Evaluation on training data (8124 cases):  
##   
## Rules   
## ----------------  
## No Errors  
##   
## 2 120( 1.5%) <<  
##   
##   
## (a) (b) <-classified as  
## ---- ----  
## 4208 (a): class edible  
## 120 3796 (b): class poisonous  
##   
##   
## Attribute usage:  
##   
## 100.00% odor  
##   
##   
## Time: 0.0 secs

mushroom\_pred <- predict(mushroom\_c5rules, mushrooms\_test)

# cross tabulation of predicted versus actual classes  
library(gmodels)  
CrossTable(mushrooms\_test$type, mushroom\_pred,  
 prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE,  
 dnn = c('actual default', 'predicted default'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1124   
##   
##   
## | predicted default   
## actual default | edible | poisonous | Row Total |   
## ---------------|-----------|-----------|-----------|  
## edible | 582 | 0 | 582 |   
## | 0.518 | 0.000 | |   
## ---------------|-----------|-----------|-----------|  
## poisonous | 15 | 527 | 542 |   
## | 0.013 | 0.469 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 597 | 527 | 1124 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Accuracy is 98.66%.