Practice-5

Harsh

22/06/2020

```
#Calling all libraries
#install.packages("C50")
#install.packages("RWeka")
#install.packages("OneR")

library(C50)
```

```
## Warning: package 'C50' was built under R version 3.6.3
```

```
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 3.6.3
```

```
library(RWeka)
```

```
## Warning: package 'RWeka' was built under R version 3.6.3
```

```
library(OneR)
```

```
## Warning: package 'OneR' was built under R version 3.6.3
```

```
##
## Attaching package: 'OneR'
```

```
## The following object is masked from 'package:RWeka':
##
##     OneR
```

Problem 1: Build an R Notebook of the bank loan decision tree example in the textbook on pages 135 to 148; the CSV file is available for download below. Show each step and add appropriate documentation. Note that the provided dataset uses values 1 and 2 in default column whereas the book has no and yes in the default column.

```
#Importing Credit data
credit_data <- read.csv("C:\\Users\\harsh\\Desktop\\Introduction to Machine learning and Data Mining\\P

#Exploring data by observing the structure and first 6 rows of the data
head(credit_data)
```

```
##   checking_balance months_loan_duration credit_history   purpose amount
## 1           < 0 DM                    6       critical  radio/tv   1169
## 2       1 - 200 DM                   48         repaid  radio/tv   5951
## 3          unknown                   12       critical education   2096
## 4           < 0 DM                   42         repaid furniture   7882
## 5           < 0 DM                   24        delayed car (new)   4870
## 6          unknown                   36         repaid education   9055
##   savings_balance employment_length installment_rate personal_status
## 1         unknown           > 7 yrs                4     single male
## 2        < 100 DM         1 - 4 yrs                2          female
## 3        < 100 DM         4 - 7 yrs                2     single male
## 4        < 100 DM         4 - 7 yrs                2     single male
## 5        < 100 DM         1 - 4 yrs                3     single male
## 6         unknown         1 - 4 yrs                2     single male
##   other_debtors residence_history                property age installment_plan
## 1          none                 4             real estate  67            none
## 2          none                 2             real estate  22            none
## 3          none                 3             real estate  49            none
## 4     guarantor                 4 building society savings  45            none
## 5          none                 4            unknown/none  53            none
## 6          none                 4            unknown/none  35            none
##    housing existing_credits default dependents telephone foreign_worker
## 1      own                2       1          1       yes            yes
## 2      own                1       2          1      none            yes
## 3      own                1       1          2      none            yes
## 4 for free                1       1          2      none            yes
## 5 for free                2       2          2      none            yes
## 6 for free                1       1          2       yes            yes
##                  job
## 1   skilled employee
## 2   skilled employee
## 3 unskilled resident
## 4   skilled employee
## 5   skilled employee
## 6 unskilled resident
```

```
str(credit_data)
```

```
## 'data.frame':    1000 obs. of  21 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","delayed",..: 1 5 1 5 2 5 5 5 5 1 ...
##  $ purpose             : Factor w/ 10 levels "business","car (new)",..: 8 8 5 6 2 5 6 3 8 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_length   : Factor w/ 5 levels "> 7 yrs","0 - 1 yrs",..: 1 3 4 4 3 3 1 3 4 5 ...
##  $ installment_rate    : int  4 2 2 2 3 2 3 2 2 4 ...
##  $ personal_status     : Factor w/ 4 levels "divorced male",..: 4 2 4 4 4 4 4 4 1 3 ...
##  $ other_debtors       : Factor w/ 3 levels "co-applicant",..: 3 3 3 2 3 3 3 3 3 3 ...
##  $ residence_history   : int  4 2 3 4 4 4 4 2 4 2 ...
##  $ property            : Factor w/ 4 levels "building society savings",..: 3 3 3 1 4 4 1 2 3 2 ...
##  $ age                 : int  67 22 49 45 53 35 53 35 61 28 ...
##  $ installment_plan    : Factor w/ 3 levels "bank","none",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ housing             : Factor w/ 3 levels "for free","own",..: 2 2 2 1 1 1 2 3 2 2 ...
```

```
##  $ existing_credits    : int  2 1 1 1 2 1 1 1 1 2 ...
##  $ default             : int  1 2 1 1 2 1 1 1 1 2 ...
##  $ dependents          : int  1 1 2 2 2 2 1 1 1 1 ...
##  $ telephone           : Factor w/ 2 levels "none","yes": 2 1 1 1 1 2 1 2 1 1 ...
##  $ foreign_worker      : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ job                 : Factor w/ 4 levels "mangement self-employed",..: 2 2 4 2 2 4 2 1 4 1 ...
```

```
#Replacing 1 and 2 with 'no' and 'yes' for default column
credit_data$default[credit_data$default == 1] <- "no"
credit_data$default[credit_data$default == 2] <- "yes"

#Converting default column to factor
credit_data$default <- as.factor(credit_data$default)

#Counting total number of checkings and savings balance
table(credit_data$checking_balance)
```

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

```
table(credit_data$savings_balance)
```

```
##
##      < 100 DM     > 1000 DM  101 - 500 DM 501 - 1000 DM       unknown
##           603            48           103            63           183
```

```
#Checking the mean min max of the loan duration and amount
summary(credit_data$months_loan_duration)
```

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

```
summary(credit_data$amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     250    1366    2320    3271    3972   18424
```

```
#Counting total number of participants who were considered as default
table(credit_data$default)
```

```
##
##  no yes
## 700 300
```

```
#Random number generation and storing random data based on the numbers generated
set.seed(12345)
credit_rand <- credit_data[order(runif(1000)), ]

#Comparing the mean min max for random data and original data
summary(credit_data$amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     250    1366    2320    3271    3972   18424
```

```
summary(credit_rand$amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     250    1366    2320    3271    3972   18424
```

```
#Comparing random data and original data
head(credit_data$amount)
```

```
## [1] 1169 5951 2096 7882 4870 9055
```

```
head(credit_rand$amount)
```

```
## [1] 1199 2576 1103 4020 1501 1568
```

```
#Creating training and testing dataset by splitting the random data
credit_train <- credit_rand[1:900, ]
credit_test <- credit_rand[901:1000, ]

#Checking the distribution of training and testing dataset
prop.table(table(credit_train$default))
```

```
##
##        no       yes
## 0.7022222 0.2977778
```

```
prop.table(table(credit_test$default))
```

```
##
##   no  yes
## 0.68 0.32
```

```
#Building the Classifier model with training data
credit_model <- C5.0(credit_train[-17], credit_train$default)

#We observe that the tree has made 57 decisions
credit_model
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = credit_train$default)
##
## Classification Tree
## Number of samples: 900
## Number of predictors: 20
##
## Tree size: 57
##
## Non-standard options: attempt to group attributes
```

```r
#Summary shows all the decisions made
summary(credit_model)
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = credit_train$default)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Jun 23 20:17:26 2020
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 900 cases (21 attributes) from undefined.data
##
## Decision tree:
##
## checking_balance = unknown: no (358/44)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...foreign_worker = no:
##     :...installment_plan in {none,stores}: no (17/1)
##     :   installment_plan = bank:
##     :   :...residence_history <= 3: yes (2)
##     :       residence_history > 3: no (2)
##     foreign_worker = yes:
##     :...credit_history in {fully repaid,
##         :                  fully repaid this bank}: yes (61/20)
##         credit_history in {critical,delayed,repaid}:
##         :...months_loan_duration <= 11: no (76/13)
##             months_loan_duration > 11:
##             :...savings_balance = > 1000 DM: no (13)
##                 savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,
##                 :                   unknown}:
##                 :...checking_balance = > 200 DM:
##                     :...dependents > 1: yes (3)
##                     :   dependents <= 1:
##                     :   :...credit_history in {delayed,repaid}: no (23/3)
##                     :       credit_history = critical:
##                     :       :...amount <= 2337: yes (3)
##                     :           amount > 2337: no (6)
##                     checking_balance = < 0 DM:
##                     :...other_debtors = guarantor:
##                     :   :...credit_history = critical: yes (1)
##                     :   :   credit_history in {delayed,repaid}: no (11/1)
##                     :   other_debtors in {co-applicant,none}:
##                     :   :...job = mangement self-employed: no (26/6)
##                     :       job in {skilled employee,unemployed non-resident,
##                     :       :       unskilled resident}:
##                     :       :...purpose in {domestic appliances,others,
##                     :           :           radio/tv,repairs,
##                     :           :           retraining}: yes (33/10)
##                     :           purpose = business:
##                     :           :...job = skilled employee: yes (3)
```

```
##                       :           :   job in {unemployed non-resident,
##                       :           :           unskilled resident}: no (3)
##                       :           purpose = education: [S1]
##                       :           purpose = car (new): [S2]
##                       :           purpose = car (used):
##                       :           :...amount > 6229: yes (5)
##                       :           :   amount <= 6229: [S3]
##                       :           purpose = furniture:
##                       :           :...months_loan_duration > 27: yes (9/1)
##                       :               months_loan_duration <= 27: [S4]
##                       checking_balance = 1 - 200 DM:
##                       :...savings_balance = unknown: no (34/6)
##                           savings_balance in {< 100 DM,101 - 500 DM,
##                           :                   501 - 1000 DM}:
##                           :...months_loan_duration > 45: yes (11/1)
##                               months_loan_duration <= 45:
##                               :...installment_plan = stores:
##                                   :...age <= 35: yes (4)
##                                   :   age > 35: no (2)
##                                   installment_plan = bank:
##                                   :...residence_history <= 1: no (3)
##                                   :   residence_history > 1:
##                                   :   :...existing_credits <= 1: yes (5)
##                                   :       existing_credits > 1:
##                                   :       :...installment_rate > 2: yes (3)
##                                   :           installment_rate <= 2: [S5]
##                                   installment_plan = none:
##                                   :...other_debtors = co-applicant: yes (3/1)
##                                       other_debtors = guarantor: no (7/1)
##                                       other_debtors = none:
##                                       :...employment_length = 4 - 7 yrs:
##                                           :...age <= 41: no (16)
##                                           :   age > 41: yes (3/1)
##                                           employment_length in {> 7 yrs,
##                                           :                     0 - 1 yrs,
##                                           :                     1 - 4 yrs,
##                                           :                     unemployed}:
##                                           :...amount > 7980: yes (7)
##                                               amount <= 7980:
##                                               :...amount > 4746: no (10)
##                                                   amount <= 4746: [S6]
##
## SubTree [S1]
##
## savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}: yes (6)
## savings_balance = unknown: no (2)
##
## SubTree [S2]
##
## savings_balance = 101 - 500 DM: no (1)
## savings_balance in {501 - 1000 DM,unknown}: yes (4)
## savings_balance = < 100 DM:
## :...personal_status in {divorced male,female,single male}: yes (29/6)
##     personal_status = married male: no (2)
```

```
##
## SubTree [S3]
##
## job in {skilled employee,unemployed non-resident}: no (8/1)
## job = unskilled resident: yes (1)
##
## SubTree [S4]
##
## employment_length in {> 7 yrs,4 - 7 yrs}: no (7/1)
## employment_length = unemployed: yes (2)
## employment_length = 0 - 1 yrs:
## :...job in {skilled employee,unemployed non-resident}: no (4)
## :   job = unskilled resident: yes (1)
## employment_length = 1 - 4 yrs:
## :...property in {building society savings,unknown/none}: no (5)
##     property in {other,real estate}:
##     :...residence_history <= 2: no (4/1)
##         residence_history > 2: yes (5)
##
## SubTree [S5]
##
## other_debtors = co-applicant: yes (1)
## other_debtors in {guarantor,none}: no (3)
##
## SubTree [S6]
##
## housing = for free: no (2)
## housing = rent:
## :...credit_history = critical: no (1)
## :   credit_history in {delayed,repaid}: yes (10/2)
## housing = own:
## :...savings_balance = 101 - 500 DM: no (6)
##     savings_balance in {< 100 DM,501 - 1000 DM}:
##     :...residence_history <= 1: no (8/1)
##         residence_history > 1:
##         :...installment_rate <= 1: no (2)
##             installment_rate > 1:
##             :...employment_length in {> 7 yrs,unemployed}: no (13/6)
##                 employment_length in {0 - 1 yrs,1 - 4 yrs}: yes (10)
##
##
## Evaluation on training data (900 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##      57  127(14.1%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     590    42    (a): class no
##      85   183    (b): class yes
```

```
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##   60.22% foreign_worker
##   57.89% credit_history
##   51.11% months_loan_duration
##   42.67% savings_balance
##   30.44% other_debtors
##   17.78% job
##   15.56% installment_plan
##   14.89% purpose
##   12.89% employment_length
##   10.22% amount
##    6.78% residence_history
##    5.78% housing
##    3.89% dependents
##    3.56% installment_rate
##    3.44% personal_status
##    2.78% age
##    1.56% property
##    1.33% existing_credits
##
##
## Time: 0.0 secs
```

```r
#Testing the accuracy of the model on the testing data
credit_pred <- predict(credit_model, credit_test)

#Calculating the accuracy. We observe that the false rate of the model is 25%
CrossTable(credit_test$default, credit_pred, prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE, dnn = c
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  100
##
##
##                | predicted default
## actual default |        no |       yes | Row Total |
## ---------------|-----------|-----------|-----------|
##             no |        54 |        14 |        68 |
##                |     0.540 |     0.140 |           |
## ---------------|-----------|-----------|-----------|
##            yes |        11 |        21 |        32 |
##                |     0.110 |     0.210 |           |
## ---------------|-----------|-----------|-----------|
```

```
##   Column Total |        65 |        35 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

```
#Improving performance by boosting method in which we set trail as 10
credit_boost10 <- C5.0(credit_train[-17], credit_train$default,trials = 10)
summary(credit_boost10)
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = credit_train$default, trials = 10)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Jun 23 20:17:26 2020
## -------------------------------
##
## Class specified by attribute 'outcome'
##
## Read 900 cases (21 attributes) from undefined.data
##
## -----  Trial 0:  -----
##
## Decision tree:
##
## checking_balance = unknown: no (358/44)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...foreign_worker = no:
##     :...installment_plan in {none,stores}: no (17/1)
##     :   installment_plan = bank:
##     :   :...residence_history <= 3: yes (2)
##     :       residence_history > 3: no (2)
##     foreign_worker = yes:
##     :...credit_history in {fully repaid,
##         :                  fully repaid this bank}: yes (61/20)
##         credit_history in {critical,delayed,repaid}:
##         :...months_loan_duration <= 11: no (76/13)
##             months_loan_duration > 11:
##             :...savings_balance = > 1000 DM: no (13)
##                 savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,
##                 :                   unknown}:
##                 :...checking_balance = > 200 DM:
##                     :...dependents > 1: yes (3)
##                     :   dependents <= 1:
##                     :   :...credit_history in {delayed,repaid}: no (23/3)
##                     :       credit_history = critical:
##                     :       :...amount <= 2337: yes (3)
##                     :           amount > 2337: no (6)
##                     checking_balance = < 0 DM:
##                     :...other_debtors = guarantor:
##                     :   :...credit_history = critical: yes (1)
##                     :   :   credit_history in {delayed,repaid}: no (11/1)
##                     :   other_debtors in {co-applicant,none}:
##                     :   :...job = mangement self-employed: no (26/6)
```

```
##                       :       job in {skilled employee,unemployed non-resident,
##                       :       :       unskilled resident}:
##                       :       :...purpose in {domestic appliances,others,
##                       :           :           radio/tv,repairs,
##                       :           :           retraining}: yes (33/10)
##                       :           purpose = business:
##                       :           :...job = skilled employee: yes (3)
##                       :           :   job in {unemployed non-resident,
##                       :           :           unskilled resident}: no (3)
##                       :           purpose = education: [S1]
##                       :           purpose = car (new): [S2]
##                       :           purpose = car (used):
##                       :           :...amount > 6229: yes (5)
##                       :           :   amount <= 6229: [S3]
##                       :           purpose = furniture:
##                       :           :...months_loan_duration > 27: yes (9/1)
##                       :               months_loan_duration <= 27: [S4]
##                       checking_balance = 1 - 200 DM:
##                       :...savings_balance = unknown: no (34/6)
##                           savings_balance in {< 100 DM,101 - 500 DM,
##                           :                   501 - 1000 DM}:
##                           :...months_loan_duration > 45: yes (11/1)
##                               months_loan_duration <= 45:
##                               :...installment_plan = stores:
##                                   :...age <= 35: yes (4)
##                                   :   age > 35: no (2)
##                                   installment_plan = bank:
##                                   :...residence_history <= 1: no (3)
##                                   :   residence_history > 1:
##                                   :   :...existing_credits <= 1: yes (5)
##                                   :       existing_credits > 1:
##                                   :       :...installment_rate > 2: yes (3)
##                                   :           installment_rate <= 2: [S5]
##                                   installment_plan = none:
##                                   :...other_debtors = co-applicant: yes (3/1)
##                                       other_debtors = guarantor: no (7/1)
##                                       other_debtors = none:
##                                       :...employment_length = 4 - 7 yrs:
##                                           :...age <= 41: no (16)
##                                           :   age > 41: yes (3/1)
##                                           employment_length in {> 7 yrs,
##                                           :                     0 - 1 yrs,
##                                           :                     1 - 4 yrs,
##                                           :                     unemployed}:
##                                           :...amount > 7980: yes (7)
##                                               amount <= 7980:
##                                               :...amount > 4746: no (10)
##                                                   amount <= 4746: [S6]
##
## SubTree [S1]
##
## savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}: yes (6)
## savings_balance = unknown: no (2)
##
```

```
## SubTree [S2]
##
## savings_balance = 101 - 500 DM: no (1)
## savings_balance in {501 - 1000 DM,unknown}: yes (4)
## savings_balance = < 100 DM:
## :...personal_status in {divorced male,female,single male}: yes (29/6)
##     personal_status = married male: no (2)
##
## SubTree [S3]
##
## job in {skilled employee,unemployed non-resident}: no (8/1)
## job = unskilled resident: yes (1)
##
## SubTree [S4]
##
## employment_length in {> 7 yrs,4 - 7 yrs}: no (7/1)
## employment_length = unemployed: yes (2)
## employment_length = 0 - 1 yrs:
## :...job in {skilled employee,unemployed non-resident}: no (4)
## :   job = unskilled resident: yes (1)
## employment_length = 1 - 4 yrs:
## :...property in {building society savings,unknown/none}: no (5)
##     property in {other,real estate}:
##     :...residence_history <= 2: no (4/1)
##         residence_history > 2: yes (5)
##
## SubTree [S5]
##
## other_debtors = co-applicant: yes (1)
## other_debtors in {guarantor,none}: no (3)
##
## SubTree [S6]
##
## housing = for free: no (2)
## housing = rent:
## :...credit_history = critical: no (1)
## :   credit_history in {delayed,repaid}: yes (10/2)
## housing = own:
## :...savings_balance = 101 - 500 DM: no (6)
##     savings_balance in {< 100 DM,501 - 1000 DM}:
##     :...residence_history <= 1: no (8/1)
##         residence_history > 1:
##         :...installment_rate <= 1: no (2)
##             installment_rate > 1:
##             :...employment_length in {> 7 yrs,unemployed}: no (13/6)
##                 employment_length in {0 - 1 yrs,1 - 4 yrs}: yes (10)
##
## -----  Trial 1:  -----
##
## Decision tree:
##
## purpose in {car (used),domestic appliances,others,radio/tv,retraining}:
## :...months_loan_duration <= 8: no (26.9)
## :   months_loan_duration > 8:
```

```
## :   :...checking_balance = unknown: no (133.6/20.4)
## :       checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :       :...installment_plan = stores: yes (13.1/5.3)
## :           installment_plan = bank:
## :           :...amount <= 10297: no (31.6/4.7)
## :           :   amount > 10297: yes (2.4)
## :           installment_plan = none:
## :           :...employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs,
## :               :                     unemployed}: no (126.2/41.1)
## :               employment_length = 0 - 1 yrs: yes (24.1/7.1)
## purpose in {business,car (new),education,furniture,repairs}:
## :...foreign_worker = no:
##     :...checking_balance in {< 0 DM,> 200 DM,unknown}: no (19.8/0.8)
##     :   checking_balance = 1 - 200 DM: yes (2.3)
##     foreign_worker = yes:
##     :...property = unknown/none:
##         :...checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}: yes (59.9/16.5)
##         :   checking_balance = unknown: no (28/11.4)
##         property in {building society savings,other,real estate}:
##         :...savings_balance in {101 - 500 DM,501 - 1000 DM}: no (56.4/16.3)
##             savings_balance = > 1000 DM:
##             :...existing_credits <= 2: no (22/0.8)
##             :   existing_credits > 2: yes (2.3)
##             savings_balance = unknown:
##             :...job in {mangement self-employed,
##             :   :       unemployed non-resident}: no (7)
##             :   job in {skilled employee,unskilled resident}:
##             :   :...purpose in {education,repairs}: no (7.1)
##             :       purpose in {business,car (new),furniture}:
##             :       :...installment_plan = bank: yes (11.5/3.9)
##             :           installment_plan = stores: no (1.6)
##             :           installment_plan = none:
##             :           :...employment_length in {> 7 yrs,4 - 7 yrs,
##             :               :                     unemployed}: no (13.4/0.8)
##             :               employment_length in {0 - 1 yrs,1 - 4 yrs}:
##             :               :...installment_rate <= 1: no (3.2)
##             :                   installment_rate > 1: yes (19.3/4)
##             savings_balance = < 100 DM:
##             :...credit_history in {delayed,
##                 :                  fully repaid this bank}: yes (38.6/10.2)
##                 credit_history in {critical,fully repaid,repaid}:
##                 :...checking_balance = > 200 DM: no (12.6/1.6)
##                     checking_balance in {< 0 DM,1 - 200 DM,unknown}:
##                     :...housing = for free: no (0.8)
##                         housing = rent:
##                         :...installment_plan in {bank,stores}: yes (6.9)
##                         :   installment_plan = none:
##                         :   :...other_debtors = co-applicant: yes (3.1)
##                         :       other_debtors = guarantor: no (0.8)
##                         :       other_debtors = none:
##                         :       :...months_loan_duration > 22: yes (13.9/1.6)
##                         :           months_loan_duration <= 22: [S1]
##                         housing = own:
##                         :...age > 60: no (9.3)
```

```
##                               age <= 60:
##                               :...existing_credits <= 1:
##                                   :...telephone = none: [S2]
##                                   :   telephone = yes: [S3]
##                                   existing_credits > 1:
##                                   :...existing_credits > 2: no (5.5)
##                                       existing_credits <= 2:
##                                       :...residence_history <= 1: no (8.6)
##                                           residence_history > 1:
##                                           :...installment_rate <= 1: no (5.4)
##                                               installment_rate > 1: [S4]
##
## SubTree [S1]
##
## credit_history = fully repaid: yes (2.4)
## credit_history in {critical,repaid}:
## :...age <= 42: no (25/4)
##     age > 42: yes (3.1)
##
## SubTree [S2]
##
## other_debtors in {co-applicant,none}: yes (59.1/18.2)
## other_debtors = guarantor: no (2.4)
##
## SubTree [S3]
##
## property in {building society savings,real estate}: no (18.8/1.6)
## property = other: yes (12.6/3.2)
##
## SubTree [S4]
##
## installment_plan in {bank,stores}: yes (16.1/2.4)
## installment_plan = none:
## :...employment_length in {> 7 yrs,0 - 1 yrs,unemployed}: yes (21.6/7.1)
##     employment_length in {1 - 4 yrs,4 - 7 yrs}: no (21.7/1.6)
##
## -----  Trial 2:  -----
##
## Decision tree:
##
## months_loan_duration > 33:
## :...age <= 26: yes (39.9/4.5)
## :   age > 26:
## :   :...checking_balance = > 200 DM: no (3.1)
## :       checking_balance = < 0 DM:
## :       :...other_debtors = co-applicant: no (2.5)
## :       :   other_debtors in {guarantor,none}: yes (31.9/10)
## :       checking_balance in {1 - 200 DM,unknown}:
## :       :...dependents > 1: no (11.5/1.3)
## :           dependents <= 1:
## :           :...personal_status = divorced male: yes (5)
## :               personal_status in {female,married male,single male}:
## :               :...property in {building society savings,
## :                   :            real estate}: no (12.4/0.6)
```

```
## :                   property in {other,unknown/none}:
## :                   :...job in {mangement self-employed,
## :                       :       unemployed non-resident}: no (28.4/9.1)
## :                       job = unskilled resident: yes (0.6)
## :                       job = skilled employee:
## :                       :...age <= 36: no (10.7/1.9)
## :                           age > 36: yes (17.4/3.2)
## months_loan_duration <= 33:
## :...credit_history = critical:
##     :...installment_plan = bank: yes (20.4/8.9)
##     :   installment_plan in {none,stores}: no (178.5/35.2)
##     credit_history in {delayed,fully repaid,fully repaid this bank,repaid}:
##     :...foreign_worker = no: no (15.6/1.9)
##         foreign_worker = yes:
##         :...personal_status = married male: no (58/19.2)
##             personal_status = divorced male:
##             :...property in {building society savings,
##             :   :            unknown/none}: yes (12.5/5)
##             :   property in {other,real estate}: no (15.1/1.3)
##             personal_status = single male:
##             :...employment_length = 0 - 1 yrs: no (29.4/7)
##             :   employment_length = unemployed: yes (13.7/3.5)
##             :   employment_length = 4 - 7 yrs:
##             :   :...other_debtors in {co-applicant,guarantor}: yes (3.1)
##             :   :   other_debtors = none: no (36.9/7.8)
##             :   employment_length = 1 - 4 yrs:
##             :   :...housing = rent: yes (12.2/3.2)
##             :   :   housing in {for free,own}:
##             :   :   :...job = unemployed non-resident: no (0)
##             :   :       job = mangement self-employed: yes (6.8/1.3)
##             :   :       job in {skilled employee,unskilled resident}:
##             :   :       :...telephone = yes: no (16.8/0.6)
##             :   :           telephone = none: [S1]
##             :   employment_length = > 7 yrs:
##             :   :...amount > 6331: yes (9.8)
##             :       amount <= 6331:
##             :       :...months_loan_duration <= 11: no (8.6)
##             :           months_loan_duration > 11:
##             :           :...installment_rate <= 2: no (7.1/3)
##             :               installment_rate > 2:
##             :               :...installment_rate <= 3: yes (8.6/1.9)
##             :                   installment_rate > 3:
##             :                   :...dependents <= 1: no (24.6/5.6)
##             :                       dependents > 1: yes (14.9/4.5)
##             personal_status = female:
##             :...credit_history = delayed: no (13.7/4.3)
##                 credit_history in {fully repaid,
##                 :                  fully repaid this bank}: yes (23.3/6.2)
##                 credit_history = repaid:
##                 :...age > 53: no (11.8)
##                     age <= 53:
##                     :...existing_credits > 1: yes (14.5/1.3)
##                         existing_credits <= 1:
##                         :...employment_length = 4 - 7 yrs: no (14.3/1.3)
```

```
##                              employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##                              :                     unemployed}:
##                              :...residence_history <= 1: no (23/5.3)
##                                  residence_history > 1:
##                                  :...installment_rate > 2: yes (58.6/15.3)
##                                      installment_rate <= 2:
##                                      :...installment_plan = stores: no (0)
##                                          installment_plan = bank: yes (2.5)
##                                          installment_plan = none:
##                                          :...dependents <= 1: no (26.1/7.4)
##                                              dependents > 1: yes (4.3/0.6)
##
## SubTree [S1]
##
## credit_history in {delayed,fully repaid this bank,repaid}: no (49.2/13.1)
## credit_history = fully repaid: yes (2.5)
##
## -----  Trial 3:  -----
##
## Decision tree:
##
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...other_debtors = guarantor: no (31.9/9.6)
## :   other_debtors in {co-applicant,none}:
## :   :...savings_balance = > 1000 DM: no (14/3.8)
## :       savings_balance = 501 - 1000 DM: yes (15.5/7)
## :       savings_balance = unknown:
## :       :...credit_history in {critical,delayed,fully repaid}: no (19.2)
## :       :   credit_history in {fully repaid this bank,repaid}:
## :       :   :...other_debtors = co-applicant: no (3)
## :       :       other_debtors = none:
## :       :       :...job in {mangement self-employed,
## :       :           :       skilled employee}: yes (39.5/14.4)
## :       :           job in {unemployed non-resident,
## :       :                   unskilled resident}: no (8.3/1.3)
## :       savings_balance = 101 - 500 DM:
## :       :...personal_status in {divorced male,female,
## :       :   :                   married male}: yes (24.2/4.2)
## :       :   personal_status = single male:
## :       :   :...other_debtors = co-applicant: yes (1.3)
## :       :       other_debtors = none:
## :       :       :...existing_credits > 3: yes (2.2)
## :       :           existing_credits <= 3:
## :       :           :...dependents <= 1: no (18/3.9)
## :       :               dependents > 1: yes (5.5/1.6)
## :       savings_balance = < 100 DM:
## :       :...job = unemployed non-resident: yes (6.3/2)
## :           job = unskilled resident:
## :           :...property in {building society savings,
## :           :   :            real estate}: no (52.5/15.3)
## :           :   property in {other,unknown/none}: yes (22.2/6.4)
## :           job = mangement self-employed:
## :           :...residence_history <= 1: no (6.6)
## :           :   residence_history > 1:
```

```
## :           :   :...checking_balance = 1 - 200 DM:
## :           :       :...personal_status in {divorced male,female,
## :           :       :   :                   married male}: yes (20.2/1.1)
## :           :       :   personal_status = single male: no (11.2/4.7)
## :           :       checking_balance = < 0 DM:
## :           :       :...installment_rate <= 1: yes (4)
## :           :           installment_rate > 1:
## :           :           :...amount <= 7166: no (23.8/4)
## :           :               amount > 7166: yes (5.6/1.1)
## :           job = skilled employee:
## :           :...installment_rate <= 2:
## :               :...foreign_worker = no: yes (2.5)
## :               :   foreign_worker = yes:
## :               :   :...age > 26: no (37.4/9.4)
## :               :       age <= 26:
## :               :       :...residence_history <= 2: yes (16.9/1.8)
## :               :           residence_history > 2: no (9.1/3.3)
## :               installment_rate > 2:
## :               :...personal_status in {divorced male,
## :                   :                   married male}: yes (21.2/7.1)
## :                   personal_status = female:
## :                   :...employment_length = > 7 yrs: no (4.7/0.5)
## :                   :   employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,
## :                   :                         unemployed}: yes (36.8/7.7)
## :                   personal_status = single male:
## :                   :...months_loan_duration <= 11: no (3.9)
## :                       months_loan_duration > 11:
## :                       :...other_debtors = co-applicant: yes (2.6)
## :                           other_debtors = none:
## :                           :...residence_history <= 1: no (4.6)
## :                               residence_history > 1: yes (48.1/11.4)
## checking_balance in {> 200 DM,unknown}:
## :...foreign_worker = no: no (9)
##     foreign_worker = yes:
##     :...installment_plan = bank:
##         :...other_debtors in {co-applicant,guarantor}: no (4.1)
##         :   other_debtors = none:
##         :   :...amount > 3711: yes (21.1/3.2)
##         :       amount <= 3711:
##         :       :...age > 41: no (10.1)
##         :           age <= 41:
##         :           :...job = mangement self-employed: no (3.4)
##         :               job in {skilled employee,unemployed non-resident,
##         :                       unskilled resident}: yes (21.4/8.6)
##         installment_plan in {none,stores}:
##         :...purpose in {car (used),domestic appliances,others,
##             :           retraining}: no (35.1)
##             purpose in {business,car (new),education,furniture,radio/tv,
##             :           repairs}:
##             :...employment_length in {> 7 yrs,4 - 7 yrs}:
##                 :...credit_history = fully repaid this bank: no (0)
##                 :   credit_history = delayed: yes (16.8/7.8)
##                 :   credit_history in {critical,fully repaid,repaid}:
##                 :   :...amount <= 8648: no (91.4/11.1)
```

```
##                 :       amount > 8648: yes (5.5/1.8)
##                 employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}:
##                 :...other_debtors in {co-applicant,guarantor}: yes (11.9/2.9)
##                     other_debtors = none:
##                     :...amount > 4153:
##                         :...job = unemployed non-resident: yes (0)
##                         :   job = unskilled resident: no (2.1)
##                         :   job in {mangement self-employed,skilled employee}:
##                         :   :...months_loan_duration <= 39: yes (30/4.5)
##                         :       months_loan_duration > 39: no (2)
##                         amount <= 4153:
##                         :...property = building society savings: no (15.4)
##                             property in {other,real estate,unknown/none}:
##                             :...checking_balance = > 200 DM:
##                                 :...months_loan_duration <= 22: yes (18.8/5.4)
##                                 :   months_loan_duration > 22: no (3.4)
##                                 checking_balance = unknown:
##                                 :...purpose in {car (new),
##                                     :           radio/tv}: no (37.2/1.5)
##                                     purpose in {business,education,furniture,
##                                     :           repairs}:
##                                     :...installment_rate <= 3: no (11.4)
##                                         installment_rate > 3: [S1]
##
## SubTree [S1]
##
## installment_plan = stores: yes (5)
## installment_plan = none:
## :...personal_status in {divorced male,single male}: no (6.2)
##     personal_status in {female,married male}: yes (15.6/2.9)
##
## -----  Trial 4:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...foreign_worker = no: no (7.3)
## :   foreign_worker = yes:
## :   :...purpose in {car (used),domestic appliances,education,furniture,others,
## :       :           retraining}: no (119.3/30.8)
## :       purpose = repairs: yes (6.4/3.2)
## :       purpose = business:
## :       :...employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs}: no (29.9/7.8)
## :       :   employment_length in {0 - 1 yrs,unemployed}: yes (11/1.2)
## :       purpose = car (new):
## :       :...installment_plan = bank: yes (13.5/4.3)
## :       :   installment_plan in {none,stores}:
## :       :   :...amount <= 11760: no (56/13.3)
## :       :       amount > 11760: yes (3)
## :       purpose = radio/tv:
## :       :...credit_history in {critical,delayed,
## :           :                  fully repaid this bank}: no (47/5.8)
## :           credit_history = fully repaid: yes (1.5/0.4)
## :           credit_history = repaid:
```

```
## :           :...existing_credits > 1: yes (11.1/2.6)
## :               existing_credits <= 1:
## :               :...age > 28: no (23.1)
## :                   age <= 28:
## :                   :...months_loan_duration <= 10: yes (6.5)
## :                       months_loan_duration > 10: no (19.5/6)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...savings_balance in {> 1000 DM,501 - 1000 DM,unknown}: no (111.6/34.6)
##     savings_balance in {< 100 DM,101 - 500 DM}:
##     :...months_loan_duration > 47: yes (31.9/5.1)
##         months_loan_duration <= 47:
##         :...purpose in {business,others,repairs,retraining}: no (51/17.6)
##             purpose in {domestic appliances,education}: yes (20.9/6.4)
##             purpose = car (used):
##             :...personal_status in {divorced male,married male,
##             :   :                   single male}: no (24.1/4.1)
##             :   personal_status = female: yes (4.9/0.9)
##             purpose = furniture:
##             :...other_debtors = guarantor: no (4.6)
##             :   other_debtors in {co-applicant,none}:
##             :   :...personal_status = married male: no (5.4)
##             :       personal_status in {divorced male,female,single male}:
##             :       :...installment_plan = stores: no (2)
##             :           installment_plan in {bank,none}:
##             :           :...housing = for free: no (5.3/1.2)
##             :               housing in {own,rent}: yes (78/28.1)
##             purpose = car (new):
##             :...other_debtors in {co-applicant,guarantor}: yes (13.9/1.5)
##             :   other_debtors = none:
##             :   :...credit_history = critical: no (23.2/5.9)
##             :       credit_history in {delayed,fully repaid,
##             :       :                  fully repaid this bank}: yes (21.2/7.1)
##             :       credit_history = repaid:
##             :       :...dependents > 1: no (8.7/2.5)
##             :           dependents <= 1:
##             :           :...months_loan_duration > 40: no (2)
##             :               months_loan_duration <= 40:
##             :               :...employment_length in {> 7 yrs,0 - 1 yrs,
##             :                   :                      1 - 4 yrs,
##             :                   :                      4 - 7 yrs}: yes (34.1/6.4)
##             :                   employment_length = unemployed: no (3)
##             purpose = radio/tv:
##             :...foreign_worker = no: no (3.3)
##                 foreign_worker = yes:
##                 :...months_loan_duration > 36: yes (6.6)
##                     months_loan_duration <= 36:
##                     :...savings_balance = 101 - 500 DM: yes (10.4/2.3)
##                         savings_balance = < 100 DM:
##                         :...other_debtors in {co-applicant,
##                             :                 guarantor}: no (13.4/2.1)
##                             other_debtors = none:
##                             :...employment_length = 4 - 7 yrs: no (5.5)
##                                 employment_length = unemployed: yes (2.3)
##                                 employment_length in {> 7 yrs,0 - 1 yrs,
```

```
##                                 :                      1 - 4 yrs}:
##                                 :...job in {mangement self-employed,
##                                     :       unemployed non-resident}: no (6.3)
##                                     job in {skilled employee,
##                                     :       unskilled resident}: [S1]
##
## SubTree [S1]
##
## personal_status in {divorced male,married male}: yes (9.2)
## personal_status in {female,single male}:
## :...housing = for free: no (0)
##     housing = rent: yes (7.6/1.2)
##     housing = own:
##     :...amount <= 918: yes (6.6/0.4)
##         amount > 918: no (27.9/6.5)
##
## -----  Trial 5:  -----
##
## Decision tree:
##
## checking_balance = unknown:
## :...employment_length in {> 7 yrs,4 - 7 yrs}:
## :   :...months_loan_duration > 24: no (28.4)
## :   :   months_loan_duration <= 24:
## :   :   :...age <= 22: yes (7.7/1.2)
## :   :       age > 22: no (86/18.7)
## :   employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}:
## :   :...months_loan_duration > 24:
## :       :...installment_rate <= 1: no (4/1)
## :       :   installment_rate > 1: yes (38.8/9.4)
## :       months_loan_duration <= 24:
## :       :...other_debtors in {co-applicant,guarantor}: yes (15.7/5.5)
## :           other_debtors = none:
## :           :...purpose in {car (new),car (used),domestic appliances,furniture,
## :               :           others,radio/tv,repairs,retraining}: no (79.2/14)
## :               purpose in {business,education}:
## :               :...amount <= 1800: no (3.4)
## :                   amount > 1800: yes (20.1/3.7)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...foreign_worker = no: no (15.5/4.3)
##     foreign_worker = yes:
##     :...credit_history = delayed:
##         :...job = unemployed non-resident: no (0)
##         :   job = mangement self-employed: yes (15.9/4.2)
##         :   job in {skilled employee,unskilled resident}:
##         :   :...checking_balance = < 0 DM: yes (11.4/4.8)
##         :       checking_balance in {> 200 DM,1 - 200 DM}: no (22.5/0.7)
##         credit_history = fully repaid:
##         :...housing in {for free,rent}: yes (7.8)
##         :   housing = own:
##         :   :...installment_rate <= 3: no (17.2/3.2)
##         :       installment_rate > 3: yes (5.5)
##         credit_history = fully repaid this bank:
##         :...other_debtors = co-applicant: no (3/0.4)
```

```
##         :   other_debtors = guarantor: yes (3.9)
##         :   other_debtors = none:
##         :   :...property in {building society savings,
##         :       :            unknown/none}: yes (19.9/2)
##         :       property in {other,real estate}: no (19/6.8)
##         credit_history = critical:
##         :...savings_balance in {> 1000 DM,501 - 1000 DM}: yes (16.3/3.6)
##         :   savings_balance in {101 - 500 DM,unknown}: no (15.2/2.9)
##         :   savings_balance = < 100 DM:
##         :   :...personal_status = divorced male: yes (9.2/1.1)
##         :       personal_status in {female,married male,single male}:
##         :       :...residence_history <= 1: no (5.7)
##         :           residence_history > 1:
##         :           :...residence_history <= 2: yes (20/6.1)
##         :               residence_history > 2: no (61.2/20.1)
##         credit_history = repaid:
##         :...amount > 8648: yes (19.9/2.8)
##             amount <= 8648:
##             :...months_loan_duration <= 8: no (18.3/2)
##                 months_loan_duration > 8:
##                 :...purpose in {business,car (used),education,
##                     :           others}: no (52.4/16.9)
##                     purpose in {domestic appliances,repairs,
##                     :           retraining}: yes (15.2/3.8)
##                     purpose = furniture:
##                     :...installment_plan = bank: yes (6.2/1.2)
##                     :   installment_plan in {none,stores}: no (68.1/27.8)
##                     purpose = car (new):
##                     :...telephone = yes: no (17.8/6)
##                     :   telephone = none:
##                     :   :...age > 32: yes (19.3/1.8)
##                     :       age <= 32:
##                     :       :...installment_rate <= 2: no (8.9/1.8)
##                     :           installment_rate > 2: yes (25.7/8.1)
##                     purpose = radio/tv:
##                     :...employment_length in {> 7 yrs,4 - 7 yrs}: no (25.4/7.6)
##                         employment_length = unemployed: yes (4.6)
##                         employment_length = 0 - 1 yrs:
##                         :...age <= 22: yes (7.3)
##                         :   age > 22: no (15.7/6)
##                         employment_length = 1 - 4 yrs:
##                         :...other_debtors = co-applicant: yes (0.8)
##                             other_debtors = guarantor: no (2.1)
##                             other_debtors = none: [S1]
##
## SubTree [S1]
##
## property in {building society savings,other}: no (18.4/6.1)
## property in {real estate,unknown/none}: yes (21.3/4.3)
##
## -----  Trial 6:  -----
##
## Decision tree:
##
```

```
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...credit_history = fully repaid: yes (29.1/11.6)
## :   credit_history = fully repaid this bank:
## :   :...age <= 23: no (3.9)
## :   :   age > 23:
## :   :   :...amount <= 409: no (3.8)
## :   :       amount > 409: yes (39/8.8)
## :   credit_history = delayed:
## :   :...installment_rate <= 1: no (9.2)
## :   :   installment_rate > 1:
## :   :   :...savings_balance in {> 1000 DM,101 - 500 DM,501 - 1000 DM,
## :   :       :                   unknown}: no (16.9/2.7)
## :   :       savings_balance = < 100 DM:
## :   :       :...months_loan_duration <= 15: no (9.1/1.6)
## :   :           months_loan_duration > 15: yes (14.6/1.7)
## :   credit_history = critical:
## :   :...other_debtors in {co-applicant,guarantor}: yes (17.2/5.5)
## :   :   other_debtors = none:
## :   :   :...savings_balance in {> 1000 DM,101 - 500 DM,unknown}: no (21.8/7.4)
## :   :       savings_balance = 501 - 1000 DM: yes (9.9/2.7)
## :   :       savings_balance = < 100 DM:
## :   :       :...dependents > 1: no (13.4/2.4)
## :   :           dependents <= 1:
## :   :           :...property in {other,real estate}:
## :   :               :...amount <= 1169: no (3.7)
## :   :               :   amount > 1169: yes (33.8/7.9)
## :   :               property in {building society savings,unknown/none}:
## :   :               :...months_loan_duration > 36: yes (4.6/0.3)
## :   :                   months_loan_duration <= 36:
## :   :                   :...amount <= 731: yes (3.3/0.9)
## :   :                       amount > 731:
## :   :                       :...amount <= 7685: no (28.4/2.3)
## :   :                           amount > 7685: yes (2.3)
## :   credit_history = repaid:
## :   :...other_debtors = guarantor: no (16.5/4.2)
## :       other_debtors in {co-applicant,none}:
## :       :...months_loan_duration <= 11:
## :           :...personal_status in {divorced male,married male,
## :           :   :                   single male}: no (24.8/2.2)
## :           :   personal_status = female:
## :           :   :...age <= 48: yes (19.5/5.3)
## :           :       age > 48: no (5.2)
## :           months_loan_duration > 11:
## :           :...foreign_worker = no: no (3.5)
## :               foreign_worker = yes:
## :               :...residence_history <= 1:
## :                   :...job in {unemployed non-resident,
## :                   :   :       unskilled resident}: yes (13.9/2.6)
## :                   :   job in {mangement self-employed,skilled employee}:
## :                   :   :...installment_rate <= 2: yes (19.9/7)
## :                   :       installment_rate > 2: no (32.8/5.5)
## :                   residence_history > 1:
## :                   :...employment_length = 0 - 1 yrs: yes (33.6/4.9)
## :                       employment_length = unemployed: no (10.9/3)
```

```
## :                       employment_length in {> 7 yrs,1 - 4 yrs}:
## :                       :...personal_status = divorced male: no (4.1)
## :                       :   personal_status in {female,married male,
## :                       :                       single male}: yes (140.3/45.8)
## :                       employment_length = 4 - 7 yrs:
## :                       :...other_debtors = co-applicant: yes (2.3)
## :                           other_debtors = none:
## :                           :...installment_rate <= 2: no (12.1/1.5)
## :                               installment_rate > 2:
## :                               :...dependents <= 1: yes (16.7/5.3)
## :                                   dependents > 1: no (5.4/1.3)
## checking_balance = unknown:
## :...installment_plan = stores: no (17.5/6.8)
##     installment_plan = bank:
##     :...employment_length in {0 - 1 yrs,4 - 7 yrs}: no (10.6/3)
##     :   employment_length in {1 - 4 yrs,unemployed}: yes (23.5/8.8)
##     :   employment_length = > 7 yrs:
##     :   :...age <= 41: yes (15.3/4.8)
##     :       age > 41: no (6.4)
##     installment_plan = none:
##     :...purpose in {car (used),domestic appliances,others,
##         :           retraining}: no (22.1)
##         purpose in {business,car (new),education,furniture,radio/tv,repairs}:
##         :...credit_history in {fully repaid,
##             :                  fully repaid this bank}: no (3.4)
##             credit_history = critical:
##             :...amount <= 6887: no (45.4/3.1)
##             :   amount > 6887: yes (5/0.3)
##             credit_history in {delayed,repaid}:
##             :...property = building society savings: yes (22.7/10.9)
##                 property = unknown/none: no (10.2/2.3)
##                 property = real estate:
##                 :...age <= 23: yes (11.1/2.4)
##                 :   age > 23: no (23.3)
##                 property = other:
##                 :...job in {mangement self-employed,unemployed non-resident,
##                     :       unskilled resident}: yes (16.3/4.3)
##                     job = skilled employee:
##                     :...residence_history <= 1: yes (3.9)
##                         residence_history > 1:
##                         :...personal_status in {divorced male,
##                             :                   female}: no (8.5/1.6)
##                             personal_status in {married male,single male}:
##                             :...housing = for free: no (0)
##                                 housing = rent: yes (4)
##                                 housing = own:
##                                 :...age <= 31: yes (17/7.4)
##                                     age > 31: no (8.1)
##
## -----  Trial 7:  -----
##
## Decision tree:
##
## checking_balance = < 0 DM:
```

```
## :...foreign_worker = no: no (10/1.9)
## :   foreign_worker = yes:
## :   :...savings_balance = > 1000 DM: no (5.7)
## :       savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,unknown}:
## :       :...credit_history in {critical,delayed}: no (68.5/31.1)
## :           credit_history in {fully repaid,
## :           :                  fully repaid this bank}: yes (36.1/11.2)
## :           credit_history = repaid:
## :           :...purpose in {business,car (used),others}: no (24.6/8.4)
## :               purpose in {domestic appliances,education,repairs,
## :               :           retraining}: yes (20/6.1)
## :               purpose = car (new):
## :               :...savings_balance = 501 - 1000 DM: yes (0)
## :               :   savings_balance = 101 - 500 DM: no (3)
## :               :   savings_balance in {< 100 DM,unknown}:
## :               :   :...other_debtors = co-applicant: yes (3.3)
## :               :       other_debtors = guarantor: no (1.2)
## :               :       other_debtors = none:
## :               :       :...installment_rate <= 2: no (9.8/2.9)
## :               :           installment_rate > 2: yes (21.6/4)
## :               purpose = furniture:
## :               :...amount > 4657: yes (5.5/0.6)
## :               :   amount <= 4657:
## :               :   :...amount > 3512: no (13/1.9)
## :               :       amount <= 3512:
## :               :       :...months_loan_duration <= 15: no (12.7/2.6)
## :               :           months_loan_duration > 15: yes (18.9/3.9)
## :               purpose = radio/tv:
## :               :...months_loan_duration > 36: yes (5.7)
## :                   months_loan_duration <= 36:
## :                   :...amount <= 909: yes (5.9)
## :                       amount > 909:
## :                       :...residence_history <= 1: no (4.7)
## :                           residence_history > 1:
## :                           :...age <= 37: yes (21.5/8.9)
## :                               age > 37: no (5.6)
## checking_balance in {> 200 DM,1 - 200 DM,unknown}:
## :...purpose in {domestic appliances,others,radio/tv,repairs,
##     :           retraining}: no (201/61)
##     purpose = car (used):
##     :...amount <= 11054: no (45.1/5.1)
##     :   amount > 11054: yes (4.3)
##     purpose = education:
##     :...savings_balance in {501 - 1000 DM,unknown}: no (6.9)
##     :   savings_balance in {< 100 DM,> 1000 DM,101 - 500 DM}:
##     :   :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##     :       :                     unemployed}: yes (29.1/7)
##     :       employment_length = 4 - 7 yrs: no (4.1/0.2)
##     purpose = business:
##     :...savings_balance in {> 1000 DM,101 - 500 DM,501 - 1000 DM,
##     :   :                   unknown}: no (30/4.4)
##     :   savings_balance = < 100 DM:
##     :   :...other_debtors in {co-applicant,guarantor}: no (2.3)
##     :       other_debtors = none:
```

```
##     :       :...employment_length = 4 - 7 yrs: no (5.9)
##     :           employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,unemployed}:
##     :           :...residence_history <= 1: no (7.9/2.2)
##     :               residence_history > 1: yes (26.9/5.4)
##     purpose = car (new):
##     :...foreign_worker = no: no (5.5)
##     :   foreign_worker = yes:
##     :   :...savings_balance = > 1000 DM: no (7)
##     :       savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,unknown}:
##     :       :...installment_plan = stores: yes (6.5/2.6)
##     :           installment_plan = bank:
##     :           :...job = mangement self-employed: no (6.6/1.9)
##     :           :   job in {skilled employee,unemployed non-resident,
##     :           :           unskilled resident}: yes (17.6/1.4)
##     :           installment_plan = none:
##     :           :...existing_credits > 2: yes (4.8/0.5)
##     :               existing_credits <= 2:
##     :               :...amount <= 12204: no (83.5/27.7)
##     :                   amount > 12204: yes (6.2)
##     purpose = furniture:
##     :...months_loan_duration > 33: yes (8.5/1.2)
##         months_loan_duration <= 33:
##         :...existing_credits > 1: no (24.3/4.2)
##             existing_credits <= 1:
##             :...credit_history in {critical,fully repaid,
##                 :                  fully repaid this bank}: no (9.6/2)
##                 credit_history = delayed: yes (3.7)
##                 credit_history = repaid:
##                 :...age <= 23: no (10.7)
##                     age > 23:
##                     :...dependents > 1: yes (4.4)
##                         dependents <= 1:
##                         :...checking_balance = > 200 DM: no (3.6)
##                             checking_balance in {1 - 200 DM,unknown}:
##                             :...months_loan_duration <= 18: yes (28.4/10.2)
##                                 months_loan_duration > 18: no (8.1/1.2)
##
## -----  Trial 8:  -----
##
## Decision tree:
##
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...savings_balance = > 1000 DM: no (17.3/4.9)
## :   savings_balance = 501 - 1000 DM: yes (20.8/10.4)
## :   savings_balance = 101 - 500 DM:
## :   :...personal_status in {divorced male,single male}: no (34.6/13.2)
## :   :   personal_status in {female,married male}: yes (22.8/4.2)
## :   savings_balance = unknown:
## :   :...installment_plan = bank: no (16.5/1.1)
## :   :   installment_plan = stores: yes (1.7)
## :   :   installment_plan = none:
## :   :   :...other_debtors in {co-applicant,guarantor}: no (3.4)
## :   :       other_debtors = none:
## :   :       :...existing_credits > 1: no (8.7/1.6)
```

```
## :   :           existing_credits <= 1:
## :   :           :...months_loan_duration <= 10: no (5.5)
## :   :               months_loan_duration > 10: yes (42/16.7)
## :   savings_balance = < 100 DM:
## :   :...months_loan_duration > 47: yes (26.9/4)
## :       months_loan_duration <= 47:
## :       :...purpose in {domestic appliances,others,retraining}: no (14.9/4.9)
## :           purpose in {education,repairs}: yes (29.8/9.1)
## :           purpose = business:
## :           :...months_loan_duration <= 18: no (8)
## :           :   months_loan_duration > 18: yes (17.3/5.1)
## :           purpose = car (used):
## :           :...residence_history <= 3: no (10/0.5)
## :           :   residence_history > 3: yes (16.9/6)
## :           purpose = car (new):
## :           :...employment_length in {> 7 yrs,0 - 1 yrs,
## :           :   :                      unemployed}: yes (55.8/11.6)
## :           :   employment_length in {1 - 4 yrs,4 - 7 yrs}:
## :           :   :...installment_plan = stores: no (0)
## :           :       installment_plan = bank: yes (7.2/1.3)
## :           :       installment_plan = none:
## :           :       :...months_loan_duration <= 22: no (27.7/6)
## :           :           months_loan_duration > 22: yes (8.4/1.7)
## :           purpose = furniture:
## :           :...other_debtors = guarantor: no (4.6)
## :           :   other_debtors in {co-applicant,none}:
## :           :   :...residence_history <= 1: no (14.8/2.9)
## :           :       residence_history > 1:
## :           :       :...age <= 36: no (51.8/24)
## :           :           age > 36: yes (26.9/6.3)
## :           purpose = radio/tv:
## :           :...months_loan_duration > 36: yes (5.7)
## :               months_loan_duration <= 36:
## :               :...other_debtors in {co-applicant,guarantor}: no (9.5/1.6)
## :                   other_debtors = none:
## :                   :...employment_length in {> 7 yrs,1 - 4 yrs,
## :                       :                      4 - 7 yrs}: no (43.8/15.6)
## :                       employment_length in {0 - 1 yrs,
## :                                             unemployed}: yes (21.4/6.8)
## checking_balance in {> 200 DM,unknown}:
## :...employment_length in {0 - 1 yrs,unemployed}:
##     :...property in {building society savings,unknown/none}: no (21.8/5)
##     :   property in {other,real estate}:
##     :   :...other_debtors = co-applicant: yes (5.5)
##     :       other_debtors = guarantor: no (0.7)
##     :       other_debtors = none:
##     :       :...amount > 4746: yes (12.8)
##     :           amount <= 4746:
##     :           :...checking_balance = > 200 DM: yes (9.3/1.6)
##     :               checking_balance = unknown: no (15.1/3.7)
##     employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs}:
##     :...months_loan_duration <= 8: no (16.3)
##         months_loan_duration > 8:
##         :...dependents > 1:
```

```
##             :...checking_balance = > 200 DM: yes (7.4)
##             :   checking_balance = unknown:
##             :   :...savings_balance in {< 100 DM,> 1000 DM}: yes (19.9/7.3)
##             :       savings_balance in {101 - 500 DM,501 - 1000 DM,
##             :                           unknown}: no (13.7)
##             dependents <= 1:
##             :...employment_length in {> 7 yrs,4 - 7 yrs}: no (87.6/15.1)
##                 employment_length = 1 - 4 yrs:
##                 :...residence_history <= 1: no (6.4)
##                     residence_history > 1:
##                     :...credit_history in {critical,delayed,fully repaid,
##                         :                  fully repaid this bank}: no (49.7/13.8)
##                         credit_history = repaid:
##                         :...installment_plan = stores: yes (2.4)
##                             installment_plan in {bank,none}:
##                             :...amount > 2569: yes (22.4/6.2)
##                                 amount <= 2569:
##                                 :...job in {mangement self-employed,
##                                     :       skilled employee,
##                                     :       unemployed non-resident}: no (17.7)
##                                     job = unskilled resident: yes (13.6/5.1)
##
## -----  Trial 9:  -----
##
## Decision tree:
##
## savings_balance in {> 1000 DM,501 - 1000 DM,unknown}: no (216.3/63.4)
## savings_balance in {< 100 DM,101 - 500 DM}:
## :...checking_balance in {> 200 DM,unknown}:
##     :...other_debtors = co-applicant: yes (11.6/4.8)
##     :   other_debtors = guarantor: no (2.7)
##     :   other_debtors = none:
##     :   :...installment_plan = stores: yes (13.7/5.2)
##     :       installment_plan = bank:
##     :       :...housing = rent: yes (5.3/0.8)
##     :       :   housing in {for free,own}:
##     :       :   :...purpose in {business,car (new)}: yes (10.4/1.5)
##     :       :       purpose in {car (used),domestic appliances,education,
##     :       :                   furniture,others,radio/tv,repairs,
##     :       :                   retraining}: no (18/1.5)
##     :       installment_plan = none:
##     :       :...credit_history in {fully repaid,repaid}: no (79.8/21.1)
##     :           credit_history = fully repaid this bank: yes (1.3)
##     :           credit_history = critical:
##     :           :...housing in {for free,own}: no (26.9)
##     :           :   housing = rent: yes (3.6/0.7)
##     :           credit_history = delayed:
##     :           :...installment_rate <= 3: no (11.5/1.7)
##     :               installment_rate > 3: yes (15.4/3.1)
##     checking_balance in {< 0 DM,1 - 200 DM}:
##     :...months_loan_duration > 47: yes (34.3/6.9)
##         months_loan_duration <= 47:
##         :...purpose in {business,domestic appliances,repairs,
##             :           retraining}: no (58.6/22.9)
```

```
##             purpose in {education,others}: yes (22.3/8.7)
##             purpose = car (used):
##             :...amount <= 9283: no (27.6/4)
##             :   amount > 9283: yes (7.6/1.3)
##             purpose = car (new):
##             :...other_debtors in {co-applicant,guarantor}: yes (14.8/1.9)
##             :   other_debtors = none:
##             :   :...foreign_worker = no: no (3.2)
##             :       foreign_worker = yes:
##             :       :...personal_status = divorced male: yes (4.5/1.7)
##             :           personal_status = married male: no (10.1/3.6)
##             :           personal_status = female:
##             :           :...amount <= 5595: yes (23.3/2.8)
##             :           :   amount > 5595: no (5.5)
##             :           personal_status = single male:
##             :           :...amount > 7685: yes (5.6)
##             :               amount <= 7685:
##             :               :...installment_rate > 3: yes (26.8/10.1)
##             :                   installment_rate <= 3: [S1]
##             purpose = furniture:
##             :...other_debtors = guarantor: no (3.8)
##             :   other_debtors in {co-applicant,none}:
##             :   :...personal_status = married male: no (4.4)
##             :       personal_status in {divorced male,female,single male}:
##             :       :...months_loan_duration > 27: yes (13.7/0.8)
##             :           months_loan_duration <= 27:
##             :           :...dependents > 1: no (4.5/0.9)
##             :               dependents <= 1: [S2]
##             purpose = radio/tv:
##             :...foreign_worker = no: no (2.8)
##                 foreign_worker = yes:
##                 :...job = unemployed non-resident: yes (0)
##                     job = mangement self-employed: no (16.7/5)
##                     job in {skilled employee,unskilled resident}:
##                     :...personal_status in {divorced male,
##                         :                   married male}: yes (15.3/1.6)
##                         personal_status in {female,single male}:
##                         :...installment_rate <= 1: no (3.8)
##                             installment_rate > 1:
##                             :...dependents > 1: yes (7.1/1.1)
##                                 dependents <= 1:
##                                 :...telephone = yes: yes (13.2/3.7)
##                                     telephone = none:
##                                     :...existing_credits <= 1: no (30.9/12.9)
##                                         existing_credits > 1: yes (7.4/2.3)
##
## SubTree [S1]
##
## credit_history in {critical,fully repaid,repaid}: no (16.8/0.4)
## credit_history in {delayed,fully repaid this bank}: yes (8.1/2.5)
##
## SubTree [S2]
##
## credit_history in {critical,delayed,fully repaid,
```

```
## :                  fully repaid this bank}: yes (33.7/12.4)
## credit_history = repaid:
## :...telephone = yes: yes (8/1.1)
##     telephone = none:
##     :...amount <= 2522: yes (21.2/5.9)
##         amount > 2522: no (17.8/2.2)
##
##
## Evaluation on training data (900 cases):
##
## Trial        Decision Tree
## -----      ----------------
##    Size      Errors
##
##    0     57  127(14.1%)
##    1     42  177(19.7%)
##    2     42  190(21.1%)
##    3     54  176(19.6%)
##    4     43  168(18.7%)
##    5     44  190(21.1%)
##    6     53  202(22.4%)
##    7     48  172(19.1%)
##    8     46  195(21.7%)
##    9     44  187(20.8%)
## boost             30( 3.3%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     629     3    (a): class no
##      27   241    (b): class yes
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##  100.00% months_loan_duration
##  100.00% purpose
##  100.00% savings_balance
##  100.00% foreign_worker
##   99.11% credit_history
##   92.22% employment_length
##   92.11% installment_plan
##   91.67% other_debtors
##   88.56% amount
##   78.78% personal_status
##   77.44% property
##   74.22% age
##   68.33% dependents
##   66.00% job
##   64.00% residence_history
##   58.44% installment_rate
##   58.00% existing_credits
##   50.22% housing
```

```
##   27.44% telephone
##
##
## Time: 0.1 secs
```

```
#Testing the boosted model on the testing data
credit_boost_pred10 <- predict(credit_boost10, credit_test)

#Calculating the accuracy of the model
CrossTable(credit_test$default, credit_boost_pred10, prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  100
##
##
##                | predicted default
## actual default |        no |       yes | Row Total |
## ---------------|-----------|-----------|-----------|
##             no |        63 |         5 |        68 |
##                |     0.630 |     0.050 |           |
## ---------------|-----------|-----------|-----------|
##            yes |        16 |        16 |        32 |
##                |     0.160 |     0.160 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        79 |        21 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

```
#The false rate is reduced from 25% to 21% for boosted model

#Cost matrix for measuring the error cost
error_cost <- matrix(c(0, 1, 4, 0), nrow = 2)

#Calculating the false rate by using cost in the function.
credit_cost <- C5.0(credit_train[-17], credit_train$default, costs = error_cost)
```

```
## Warning: no dimnames were given for the cost matrix; the factor levels will be
## used
```

```
credit_cost_pred <- predict(credit_cost, credit_test)
CrossTable(credit_test$default, credit_cost_pred, prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE, d
```

```
##
```

```
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  100
##
##
##                | predicted default
## actual default |        no |       yes | Row Total |
## ---------------|-----------|-----------|-----------|
##             no |        38 |        30 |        68 |
##                |     0.380 |     0.300 |           |
## ---------------|-----------|-----------|-----------|
##            yes |         5 |        27 |        32 |
##                |     0.050 |     0.270 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        43 |        57 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

```
#We recieve false rate as 35%
```

Problem 2: Build and R Notebook of the poisonous mushrooms example using rule learners in the textbook on pages 160 to 168. Show each step and add appropriate documentation. The CSV file is available below. If you have issues with the RWeka package on MacOS, consider using a Windows computer, RStudio.cloud or skip this question.

```
#Importing mushroom data using read.csv() function
mushroom_data <- read.csv("C:\\Users\\harsh\\Desktop\\Introduction to Machine learning and Data Mining\\

#Exploring mushroom data
str(mushroom_data)
```

```
## '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 ...
```

```
#Since veil_type provides no meaningful information we remove it
mushroom_data$veil_type <- NULL

#Counting types of mushromms
table(mushroom_data$type)
```

```
##
##    edible poisonous
##      4208      3916
```

```
#Using OneR() rule learner to classify the mushroom
mushroom_1R <- OneR(type ~ ., data = mushroom_data)
mushroom_1R
```

```
##
## Call:
## OneR.formula(formula = type ~ ., data = mushroom_data)
##
## Rules:
## If odor = almond   then type = edible
## If odor = anise    then type = edible
## If odor = creosote then type = poisonous
## If odor = fishy    then type = poisonous
## If odor = foul     then type = poisonous
## If odor = musty    then type = poisonous
## If odor = none     then type = edible
## If odor = pungent  then type = poisonous
## If odor = spicy    then type = poisonous
##
## Accuracy:
## 8004 of 8124 instances classified correctly (98.52%)
```

```
#Observing the accuracy of the model
summary(mushroom_1R)
```

```
##
## Call:
## OneR.formula(formula = type ~ ., data = mushroom_data)
##
## Rules:
## If odor = almond   then type = edible
## If odor = anise    then type = edible
```

```
## If odor = creosote then type = poisonous
## If odor = fishy    then type = poisonous
## If odor = foul     then type = poisonous
## If odor = musty    then type = poisonous
## If odor = none     then type = edible
## If odor = pungent  then type = poisonous
## If odor = spicy    then type = poisonous
##
## Accuracy:
## 8004 of 8124 instances classified correctly (98.52%)
##
## Contingency table:
##             odor
## type         almond anise creosote fishy   foul musty   none pungent spicy  Sum
##   edible     * 400 * 400        0     0      0     0 * 3408       0     0 4208
##   poisonous      0     0    * 192 * 576 * 2160  * 36    120   * 256 * 576 3916
##   Sum          400   400      192   576   2160    36   3528     256   576 8124
## ---
## Maximum in each column: '*'
##
## Pearson's Chi-squared test:
## X-squared = 7659.7, df = 8, p-value < 2.2e-16
```

```
#Using Ripper algorithm to classify the mushroom type
mushroom_JRip <- JRip(type ~ ., data = mushroom_data)
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
```

```
#Ripper algorithm is better compared to OneR as it considers many features.
```

Problem 3: So far we have explored four different approaches to classification: kNN, Naive Bayes, C5.0 Decision Trees, and RIPPER Rules. Comment on the differences of the algorithms and when each is generally used. Provide examples of when they work well and when they do not work well. Add your comments to your R Notebook. Be specific and explicit; however, no code examples are needed.

kNN: 1. KNN is a non-parametric model and supports non-linear solutions. 2. It is easy to implement but is quite slow. Large computation cost during runtime if sample size is large. Because of which it is known as lazy learning algorithm. 3. Usually Euclidean distance is used to calculate distances. Manhattan distance, Hamming Distance, Minkowski distance are different alternatives. 4. Two types of rescaling methods can

be used for kNN which are min-max normalization and z-score normalization. 5. It can be used as both regression as well as classification. Class package is used to implement kNN.

Naive Bayes: 1. Naive bayes is parametric. And compared to kNN it is faster. 2. It is based on Naive Bayes probabilistic approach. 3. Most common application is text classification. 4. It makes use of frequency tables for each and every word with the help of document2matrix function. 5. Laplace estimator helps in reducing the error in classification as it assigns one additional count to frequency table which makes each feature non-zero. 6. corpus function is used to remove unwanted characters from the document.

C5.0 Decision Trees: 1. C5.0 decision trees makes use of the features to create new decisions. It follows divide and conquer approach. 2. It uses only the most important features from the dataset. 3. C5.0 decision tree models are often biased toward splits on features having a large number of levels 4. One of the disadvantage is that trees can continue to grow indefinitely, choosing splitting features and dividing into smaller and smaller partitions which makes it harder to interpret. 5. C5.0 uses entropy for measuring purity.

RIPPER Rules: 1. Rule learners are generally applied to problems where the features are primarily or entirely nominal 2. It is efficient for large and noisy datasets 3. Compared to decision trees, rule learners create simpler models. 4. It doesn't work with numeric data. Features have to be categorical. 5. Rule learners like RIPPER, separate-and-conquer data to identify logical if-else rules.

Problem 4: Much of our focus so far has been on building a single model that is most accurate. In practice, data scientists often construct multiple models and then combine them into a single prediction model. This is referred to as a model ensemble. Two common techniques for assembling such models are boosting and bagging. Do some research and define what model ensembles are, why they are important, and how boosting and bagging function in the construction of assemble models. Be detailed and provide references to your research. You can use this excerpt from Kelleher, MacNamee, and D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics as a starting point. This book is an excellent resource for those who want to dig deeper into data mining and machine learning.

Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging) and bias (boosting).

Boosting : 1. Boosting is used to increase performance by adding more weak learners. 2. It uses ensembles of models trained on resampled data and a vote to determine the final prediction. 3. In Boosting, each tree attempts to minimize the errors of previous tree. 4. Every new subsets contains the elements that were misclassified by previous models. 5. Sometimes, it tends to over-fit a model. 6. In some test cases it is proven to be better than bagging. 7. Example of boosting is gradient boosting.

Bagging : 1. Bagging is used when our goal is to reduce the variance of a decision tree. 2. It consists of each model in the ensemble vote with equal weight. 3. Multiple subsets are created from the original dataset, selecting observations with replacement. A weak model is created on each of these subsets. 4. Each model is trained individually, and combined using an averaging process. 5. For Classification either the most voted class is accepted (hard-voting), or the highest average of all the class probabilities is taken as the output (soft-voting). 6. Bagging is used when we have an over-fitting problem for a single model. 7. Example of bagging is random forest.