DA5030.A5.Parpattedar

Shruti Parpattedar

February 24, 2019

Question 1

Step 1 – Collecting data

Downloading the credit dataset

Step 2 - Exploring and preparing the data

Exploring the dataset and displaying the various statistics of a few columns

```
credit <- read.csv("credit.csv")
credit$default <- factor(credit$default)
str(credit)
```

```
## '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             : Factor w/ 2 levels "1","2": 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 ...
```

```
table(credit$checking_balance)
```

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

```
table(credit$savings_balance)
```

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

```
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(credit$amount)
```

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

```
table(credit$default)
```

```
##
##   1   2
## 700 300
```

Data preparation – Creating random training and test datasets

Creating training and testing datatsets. Checking if the data is proportionally distributed between the two sets.

```
set.seed(123)
train_sample <- sample(1000, 900)
str(train_sample)
```

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

```
credit_train <- credit[train_sample, ]
credit_test <- credit[-train_sample, ]
prop.table(table(credit_train$default))
```

```
##
##         1         2
## 0.7033333 0.2966667
```

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

```
##
##    1    2
## 0.67 0.33
```

Step 3 – training a model on the data

Fitting classification tree models using Quinlan's C5.0 algorithm to the training dataset

```
library(C50)
credit_model <- C5.0(credit_train[-17], credit_train$default)
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: 54
##
```

```
## Non-standard options: attempt to group attributes
```

```
summary(credit_model)
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = credit_train$default)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Feb 26 23:39:00 2019
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 900 cases (21 attributes) from undefined.data
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}: 1 (412/50)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...other_debtors = guarantor:
##     :...months_loan_duration > 36: 2 (4/1)
##     :   months_loan_duration <= 36:
##     :   :...installment_plan in {none,stores}: 1 (24)
##     :       installment_plan = bank:
##     :       :...purpose = car (new): 2 (3)
##     :           purpose in {business,car (used),domestic appliances,education,
##     :                       furniture,others,radio/tv,repairs,
##     :                       retraining}: 1 (7/1)
##     other_debtors in {co-applicant,none}:
##     :...credit_history = critical: 1 (102/30)
##         credit_history = fully repaid: 2 (27/6)
##         credit_history = fully repaid this bank:
##         :...other_debtors = co-applicant: 1 (2)
##         :   other_debtors = none: 2 (26/8)
##         credit_history in {delayed,repaid}:
##         :...savings_balance in {> 1000 DM,501 - 1000 DM}: 1 (19/3)
##             savings_balance = 101 - 500 DM:
##             :...other_debtors = co-applicant: 2 (3)
##             :   other_debtors = none:
##             :   :...personal_status in {divorced male,
##             :       :                   married male}: 2 (6/1)
##             :       personal_status = female:
##             :       :...installment_rate <= 3: 1 (4/1)
##             :       :   installment_rate > 3: 2 (4)
##             :       personal_status = single male:
##             :       :...age <= 41: 1 (15/2)
##             :           age > 41: 2 (2)
##             savings_balance = unknown:
##             :...credit_history = delayed: 1 (8)
##             :   credit_history = repaid:
##             :   :...foreign_worker = no: 1 (2)
##             :       foreign_worker = yes:
##             :       :...checking_balance = < 0 DM:
##             :           :...telephone = none: 2 (11/2)
```

```
##             :           :   telephone = yes:
##             :           :   :...amount <= 5045: 1 (5/1)
##             :           :       amount > 5045: 2 (2)
##             :           checking_balance = 1 - 200 DM:
##             :           :...residence_history > 3: 1 (9)
##             :               residence_history <= 3: [S1]
##             savings_balance = < 100 DM:
##             :...months_loan_duration > 39:
##                 :...residence_history <= 1: 1 (2)
##                 :   residence_history > 1: 2 (19/1)
##                 months_loan_duration <= 39:
##                 :...purpose in {car (new),retraining}: 2 (47/16)
##                     purpose in {domestic appliances,others}: 1 (3)
##                     purpose = car (used):
##                     :...amount <= 8086: 1 (9/1)
##                     :   amount > 8086: 2 (5)
##                     purpose = education:
##                     :...checking_balance = < 0 DM: 2 (5)
##                     :   checking_balance = 1 - 200 DM: 1 (2)
##                     purpose = repairs:
##                     :...residence_history <= 3: 2 (4/1)
##                     :   residence_history > 3: 1 (3)
##                     purpose = business:
##                     :...credit_history = delayed: 2 (2)
##                     :   credit_history = repaid:
##                     :   :...age <= 34: 1 (5)
##                     :       age > 34: 2 (2)
##                     purpose = radio/tv:
##                     :...employment_length in {0 - 1 yrs,
##                     :   :                      unemployed}: 2 (14/5)
##                     :   employment_length = 4 - 7 yrs: 1 (3)
##                     :   employment_length = > 7 yrs:
##                     :   :...amount <= 932: 2 (2)
##                     :   :   amount > 932: 1 (7)
##                     :   employment_length = 1 - 4 yrs:
##                     :   :...months_loan_duration <= 15: 1 (6)
##                     :       months_loan_duration > 15:
##                     :       :...amount <= 3275: 2 (7)
##                     :           amount > 3275: 1 (2)
##                     purpose = furniture:
##                     :...residence_history <= 1: 1 (8/1)
##                         residence_history > 1:
##                         :...installment_plan in {bank,stores}: 1 (3/1)
##                             installment_plan = none:
##                             :...telephone = yes: 2 (7/1)
##                                 telephone = none:
##                                 :...months_loan_duration > 27: 2 (3)
##                                     months_loan_duration <= 27: [S2]
##
## SubTree [S1]
##
## property in {building society savings,unknown/none}: 2 (4)
## property = other: 1 (6)
## property = real estate:
```

```
## :...job = skilled employee: 2 (2)
##     job in {mangement self-employed,unemployed non-resident,
##             unskilled resident}: 1 (2)
##
## SubTree [S2]
##
## checking_balance = 1 - 200 DM: 2 (5/2)
## checking_balance = < 0 DM:
## :...property in {building society savings,real estate,unknown/none}: 1 (8)
##     property = other:
##     :...installment_rate <= 1: 1 (2)
##         installment_rate > 1: 2 (4)
##
##
## Evaluation on training data (900 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##      54  135(15.0%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     589    44    (a): class 1
##      91   176    (b): class 2
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##   54.22% other_debtors
##   50.00% credit_history
##   32.56% savings_balance
##   25.22% months_loan_duration
##   19.78% purpose
##   10.11% residence_history
##    7.33% installment_plan
##    5.22% telephone
##    4.78% foreign_worker
##    4.56% employment_length
##    4.33% amount
##    3.44% personal_status
##    3.11% property
##    2.67% age
##    1.56% installment_rate
##    0.44% job
##
##
## Time: 0.0 secs
```

Step 4 – evaluating model performance

Using the classifer to make predictions, using the test dataset Results show that the model has an error rate of(19+7) = 26%

```
library(gmodels)
credit_pred <- predict(credit_model, credit_test)
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 |         1 |         2 | Row Total |
## ---------------|-----------|-----------|-----------|
##              1 |        60 |         7 |        67 |
##                |     0.600 |     0.070 |           |
## ---------------|-----------|-----------|-----------|
##              2 |        19 |        14 |        33 |
##                |     0.190 |     0.140 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        79 |        21 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

Step 5 – Improving model performance

Using adaptive boosting to improve model performance Although this model resulted in a higher error rate of 41 percent, it has resulted in a steep decrease in the false negatives at the expense of false positives

```
# Using 10 trials
credit_boost10 <- C5.0(credit_train[-17], credit_train$default, trials = 10)

# We observe the tree size has shrunk form 54 to 49.7
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: 20
##
```

```
## Number of boosting iterations: 10
## Average tree size: 49.7
##
## Non-standard options: attempt to group attributes
```

```
# Viewing model performance
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 Feb 26 23:39:00 2019
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 900 cases (21 attributes) from undefined.data
##
## -----  Trial 0:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}: 1 (412/50)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...other_debtors = guarantor:
##     :...months_loan_duration > 36: 2 (4/1)
##     :   months_loan_duration <= 36:
##     :   :...installment_plan in {none,stores}: 1 (24)
##     :       installment_plan = bank:
##     :       :...purpose = car (new): 2 (3)
##     :           purpose in {business,car (used),domestic appliances,education,
##     :                       furniture,others,radio/tv,repairs,
##     :                       retraining}: 1 (7/1)
##     other_debtors in {co-applicant,none}:
##     :...credit_history = critical: 1 (102/30)
##         credit_history = fully repaid: 2 (27/6)
##         credit_history = fully repaid this bank:
##         :...other_debtors = co-applicant: 1 (2)
##         :   other_debtors = none: 2 (26/8)
##         credit_history in {delayed,repaid}:
##         :...savings_balance in {> 1000 DM,501 - 1000 DM}: 1 (19/3)
##             savings_balance = 101 - 500 DM:
##             :...other_debtors = co-applicant: 2 (3)
##             :   other_debtors = none:
##             :   :...personal_status in {divorced male,
##             :       :                   married male}: 2 (6/1)
##             :       personal_status = female:
##             :       :...installment_rate <= 3: 1 (4/1)
##             :       :   installment_rate > 3: 2 (4)
##             :       personal_status = single male:
##             :       :...age <= 41: 1 (15/2)
##             :           age > 41: 2 (2)
##             savings_balance = unknown:
```

```
##             :...credit_history = delayed: 1 (8)
##             :   credit_history = repaid:
##             :   :...foreign_worker = no: 1 (2)
##             :       foreign_worker = yes:
##             :       :...checking_balance = < 0 DM:
##             :           :...telephone = none: 2 (11/2)
##             :           :   telephone = yes:
##             :           :   :...amount <= 5045: 1 (5/1)
##             :           :       amount > 5045: 2 (2)
##             :           checking_balance = 1 - 200 DM:
##             :           :...residence_history > 3: 1 (9)
##             :               residence_history <= 3: [S1]
##             savings_balance = < 100 DM:
##             :...months_loan_duration > 39:
##                 :...residence_history <= 1: 1 (2)
##                 :   residence_history > 1: 2 (19/1)
##                 months_loan_duration <= 39:
##                 :...purpose in {car (new),retraining}: 2 (47/16)
##                     purpose in {domestic appliances,others}: 1 (3)
##                     purpose = car (used):
##                     :...amount <= 8086: 1 (9/1)
##                     :   amount > 8086: 2 (5)
##                     purpose = education:
##                     :...checking_balance = < 0 DM: 2 (5)
##                     :   checking_balance = 1 - 200 DM: 1 (2)
##                     purpose = repairs:
##                     :...residence_history <= 3: 2 (4/1)
##                     :   residence_history > 3: 1 (3)
##                     purpose = business:
##                     :...credit_history = delayed: 2 (2)
##                     :   credit_history = repaid:
##                     :   :...age <= 34: 1 (5)
##                     :       age > 34: 2 (2)
##                     purpose = radio/tv:
##                     :...employment_length in {0 - 1 yrs,
##                     :   :                      unemployed}: 2 (14/5)
##                     :   employment_length = 4 - 7 yrs: 1 (3)
##                     :   employment_length = > 7 yrs:
##                     :   :...amount <= 932: 2 (2)
##                     :   :   amount > 932: 1 (7)
##                     :   employment_length = 1 - 4 yrs:
##                     :   :...months_loan_duration <= 15: 1 (6)
##                     :       months_loan_duration > 15:
##                     :       :...amount <= 3275: 2 (7)
##                     :           amount > 3275: 1 (2)
##                     purpose = furniture:
##                     :...residence_history <= 1: 1 (8/1)
##                         residence_history > 1:
##                         :...installment_plan in {bank,stores}: 1 (3/1)
##                             installment_plan = none:
##                             :...telephone = yes: 2 (7/1)
##                                 telephone = none:
##                                 :...months_loan_duration > 27: 2 (3)
##                                     months_loan_duration <= 27: [S2]
```

```
##
## SubTree [S1]
##
## property in {building society savings,unknown/none}: 2 (4)
## property = other: 1 (6)
## property = real estate:
## :...job = skilled employee: 2 (2)
##     job in {mangement self-employed,unemployed non-resident,
##             unskilled resident}: 1 (2)
##
## SubTree [S2]
##
## checking_balance = 1 - 200 DM: 2 (5/2)
## checking_balance = < 0 DM:
## :...property in {building society savings,real estate,unknown/none}: 1 (8)
##     property = other:
##     :...installment_rate <= 1: 1 (2)
##         installment_rate > 1: 2 (4)
##
## -----  Trial 1:  -----
##
## Decision tree:
##
## foreign_worker = no: 1 (28.4/2.4)
## foreign_worker = yes:
## :...checking_balance = unknown:
##     :...installment_plan in {bank,stores}:
##     :   :...other_debtors in {co-applicant,guarantor}: 1 (2.4)
##     :   :   other_debtors = none:
##     :   :   :...employment_length in {> 7 yrs,0 - 1 yrs,
##     :   :       :                     4 - 7 yrs}: 1 (32.3/10.8)
##     :   :       employment_length in {1 - 4 yrs,unemployed}: 2 (31/7.1)
##     :   installment_plan = none:
##     :   :...credit_history in {critical,fully repaid,fully repaid this bank,
##     :       :                  repaid}: 1 (224.7/32.5)
##     :       credit_history = delayed:
##     :       :...residence_history <= 1: 2 (4.3)
##     :           residence_history > 1:
##     :           :...installment_rate <= 3: 1 (11.9)
##     :               installment_rate > 3: 2 (14.2/5.6)
##     checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
##     :...other_debtors = co-applicant: 2 (24.3/7.9)
##         other_debtors = guarantor:
##         :...property in {building society savings,real estate,
##         :   :            unknown/none}: 1 (27.6/4)
##         :   property = other: 2 (3)
##         other_debtors = none:
##         :...installment_rate <= 2:
##             :...purpose in {business,car (new),car (used),domestic appliances,
##             :   :           others,radio/tv,retraining}: 1 (125.5/34.3)
##             :   purpose in {education,repairs}: 2 (13.6/4.8)
##             :   purpose = furniture:
##             :   :...job in {mangement self-employed,
##             :       :       unemployed non-resident}: 2 (4.3)
```

```
##             :       job in {skilled employee,unskilled resident}:
##             :       :...dependents > 1: 2 (2.2)
##             :           dependents <= 1:
##             :           :...checking_balance = > 200 DM: 1 (4)
##             :               checking_balance in {< 0 DM,1 - 200 DM}:
##             :               :...telephone = none: 2 (24.9/10.1)
##             :                   telephone = yes: 1 (10.1/2.4)
##             installment_rate > 2:
##             :...residence_history <= 1: 1 (39/8.5)
##                 residence_history > 1:
##                 :...credit_history = fully repaid: 2 (11.7)
##                     credit_history in {critical,delayed,fully repaid this bank,
##                     :                  repaid}:
##                     :...months_loan_duration <= 11:
##                         :...purpose in {business,car (new),car (used),
##                         :   :           domestic appliances,furniture,others,
##                         :   :           radio/tv,repairs,
##                         :   :           retraining}: 1 (35.2/6.9)
##                         :   purpose = education: 2 (5.3/0.8)
##                         months_loan_duration > 11:
##                         :...savings_balance = > 1000 DM: 1 (9.1/2.2)
##                             savings_balance = 501 - 1000 DM: 2 (15.4/5.9)
##                             savings_balance = 101 - 500 DM:
##                             :...installment_plan in {bank,stores}: 2 (8.3/0.8)
##                             :   installment_plan = none: 1 (16.2/4.5)
##                             savings_balance = unknown:
##                             :...checking_balance in {< 0 DM,
##                             :   :                    > 200 DM}: 2 (20.8/5.6)
##                             :   checking_balance = 1 - 200 DM: 1 (12.7/1.6)
##                             savings_balance = < 100 DM:
##                             :...installment_plan in {bank,
##                                 :                    stores}: 2 (25.3/3.2)
##                                 installment_plan = none:
##                                 :...dependents > 1: 1 (14.4/5.6)
##                                     dependents <= 1:
##                                     :...months_loan_duration > 42: 2 (11.5)
##                                         months_loan_duration <= 42: [S1]
##
## SubTree [S1]
##
## credit_history in {delayed,fully repaid this bank}: 2 (5.3)
## credit_history = repaid:
## :...job in {mangement self-employed,unskilled resident}: 1 (23.2/8.7)
## :   job in {skilled employee,unemployed non-resident}: 2 (24.2/7.1)
## credit_history = critical:
## :...existing_credits <= 1: 1 (6.9/2.2)
##     existing_credits > 1:
##     :...purpose in {business,car (new),domestic appliances,education,furniture,
##         :           others,repairs,retraining}: 2 (22.7/3.2)
##         purpose in {car (used),radio/tv}: 1 (4)
##
## -----  Trial 2:  -----
##
## Decision tree:
```

```
##
## checking_balance = unknown:
## :...installment_plan = bank:
## :   :...other_debtors = guarantor: 2 (0)
## :   :   other_debtors = co-applicant: 1 (1.3)
## :   :   other_debtors = none:
## :   :   :...months_loan_duration <= 8: 1 (3.4)
## :   :       months_loan_duration > 8: 2 (44.9/16.4)
## :   installment_plan in {none,stores}:
## :   :...employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs}:
## :       :...installment_rate <= 3: 1 (91.9/5.8)
## :       :   installment_rate > 3:
## :       :   :...age > 30: 1 (70.1/5.3)
## :       :       age <= 30:
## :       :       :...other_debtors = co-applicant: 1 (0.6)
## :       :           other_debtors = guarantor: 2 (3.5/0.6)
## :       :           other_debtors = none:
## :       :           :...housing = for free: 1 (0.6)
## :       :               housing = rent: 2 (4.8/1.9)
## :       :               housing = own:
## :       :               :...amount <= 1445: 1 (8)
## :       :                   amount > 1445: 2 (23.7/8)
## :       employment_length in {0 - 1 yrs,unemployed}:
## :       :...other_debtors = guarantor: 1 (0)
## :           other_debtors = co-applicant: 2 (8.6)
## :           other_debtors = none:
## :           :...months_loan_duration > 30: 2 (7.5)
## :               months_loan_duration <= 30:
## :               :...housing in {for free,rent}: 1 (5.8)
## :                   housing = own:
## :                   :...amount > 4594: 2 (5.8)
## :                       amount <= 4594:
## :                       :...purpose in {business,repairs}: 2 (4.6)
## :                           purpose in {car (new),car (used),
## :                                       domestic appliances,education,
## :                                       furniture,others,radio/tv,
## :                                       retraining}: 1 (20.7)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...months_loan_duration > 42:
##     :...savings_balance in {< 100 DM,> 1000 DM,101 - 500 DM}: 2 (42.1/6.1)
##     :   savings_balance in {501 - 1000 DM,unknown}: 1 (7.2)
##     months_loan_duration <= 42:
##     :...foreign_worker = no: 1 (15.8/3)
##         foreign_worker = yes:
##         :...other_debtors = co-applicant: 1 (26.3/12.7)
##             other_debtors = guarantor:
##             :...installment_plan = bank: 2 (9.5/3.2)
##             :   installment_plan in {none,stores}: 1 (17.5/1.5)
##             other_debtors = none:
##             :...purpose in {domestic appliances,others,
##                 :           retraining}: 1 (10/1.9)
##                 purpose = repairs: 2 (14.2/6.1)
##                 purpose = education:
##                 :...checking_balance = < 0 DM: 2 (10.1)
```

```
##                 :   checking_balance in {> 200 DM,1 - 200 DM}: 1 (18.2/7.3)
##                purpose = business:
##                :...months_loan_duration <= 18: 1 (11.3)
##                :   months_loan_duration > 18:
##                :   :...telephone = none: 1 (10.4/2.8)
##                :       telephone = yes: 2 (19.9/6)
##                purpose = car (used):
##                :...credit_history in {critical,delayed,
##                :   :                  fully repaid}: 1 (7.8)
##                :   credit_history in {fully repaid this bank,repaid}:
##                :   :...amount <= 3161: 1 (6.5)
##                :       amount > 3161: 2 (20.4/5.7)
##                purpose = car (new):
##                :...credit_history = delayed: 1 (14.6/6.7)
##                :   credit_history in {fully repaid,
##                :   :                  fully repaid this bank}: 2 (11/1.8)
##                :   credit_history = critical:
##                :   :...installment_rate <= 3: 1 (9.3)
##                :   :   installment_rate > 3: 2 (21/6.9)
##                :   credit_history = repaid:
##                :   :...personal_status = divorced male: 2 (3)
##                :       personal_status = married male: 1 (6.3/2.2)
##                :       personal_status = female:
##                :       :...job in {mangement self-employed,
##                :       :   :       unemployed non-resident}: 1 (2.6)
##                :       :   job in {skilled employee,
##                :       :           unskilled resident}: 2 (27.2/3.5)
##                :       personal_status = single male:
##                :       :...amount <= 8229: 1 (29.5/9.1)
##                :           amount > 8229: 2 (6)
##                purpose = radio/tv:
##                :...employment_length in {> 7 yrs,4 - 7 yrs}: 1 (34.3/5)
##                :   employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}:
##                :   :...existing_credits > 1: 2 (13.6/2.2)
##                :       existing_credits <= 1:
##                :       :...savings_balance in {> 1000 DM,101 - 500 DM,
##                :           :                   unknown}: 2 (7.3/1.3)
##                :           savings_balance = 501 - 1000 DM: 1 (6.5/1.8)
##                :           savings_balance = < 100 DM:
##                :           :...amount > 4473: 1 (4.2)
##                :               amount <= 4473:
##                :               :...months_loan_duration <= 7: 1 (2.4)
##                :                   months_loan_duration > 7: 2 (40.6/11.5)
##                purpose = furniture:
##                :...installment_plan = stores: 1 (11.2)
##                    installment_plan in {bank,none}:
##                    :...dependents > 1: 2 (5.2/0.6)
##                        dependents <= 1:
##                        :...checking_balance = > 200 DM: 1 (6.9)
##                            checking_balance in {< 0 DM,1 - 200 DM}:
##                            :...savings_balance in {> 1000 DM,
##                                :                   unknown}: 1 (14/4.3)
##                                savings_balance in {101 - 500 DM,
##                                :                   501 - 1000 DM}: 2 (3.7/0.6)
```

```
##                                 savings_balance = < 100 DM:
##                                 :...job in {mangement self-employed,
##                                     :       unemployed non-resident,
##                                     :       unskilled resident}: 2 (24.6/9.1)
##                                     job = skilled employee: [S1]
##
## SubTree [S1]
##
## credit_history in {critical,delayed,fully repaid,repaid}: 1 (38.6/13.8)
## credit_history = fully repaid this bank: 2 (2.8)
##
## -----  Trial 3:  -----
##
## Decision tree:
##
## checking_balance = unknown:
## :...employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs}: 1 (235.6/50.4)
## :   employment_length in {0 - 1 yrs,unemployed}:
## :   :...other_debtors = guarantor: 1 (0)
## :       other_debtors = co-applicant: 2 (7.5/0.5)
## :       other_debtors = none:
## :       :...purpose = others: 1 (0)
## :           purpose in {business,repairs}: 2 (9)
## :           purpose in {car (new),car (used),domestic appliances,education,
## :           :           furniture,radio/tv,retraining}:
## :           :...amount <= 4594: 1 (23.4)
## :               amount > 4594: 2 (11.8/1.1)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...other_debtors = guarantor: 1 (31.5/9.1)
##     other_debtors = co-applicant:
##     :...savings_balance in {> 1000 DM,501 - 1000 DM}: 2 (0)
##     :   savings_balance = unknown: 1 (3.5)
##     :   savings_balance in {< 100 DM,101 - 500 DM}:
##     :   :...amount <= 2022: 1 (5.4)
##     :       amount > 2022:
##     :       :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##     :           :                     4 - 7 yrs}: 2 (24.5/2.4)
##     :           employment_length = unemployed: 1 (2.4)
##     other_debtors = none:
##     :...purpose in {domestic appliances,others}: 2 (9.8/4.6)
##         purpose in {repairs,retraining}: 1 (22/8)
##         purpose = car (used):
##         :...personal_status in {divorced male,single male}: 1 (29.7/6.9)
##         :   personal_status in {female,married male}: 2 (13/4.1)
##         purpose = education:
##         :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##         :   :                     unemployed}: 2 (25.7/5.9)
##         :   employment_length = 4 - 7 yrs: 1 (5.9/1.4)
##         purpose = business:
##         :...age > 46: 2 (5.2)
##         :   age <= 46:
##         :   :...amount <= 10722: 1 (43.7/12.9)
##         :       amount > 10722: 2 (3.7)
##         purpose = car (new):
```

```
##         :...credit_history = critical:
##         :   :...personal_status in {divorced male,female,
##         :   :   :                   single male}: 1 (31.7/7.2)
##         :   :   personal_status = married male: 2 (4.3)
##         :   credit_history in {delayed,fully repaid,fully repaid this bank,
##         :   :                  repaid}:
##         :   :...installment_rate > 2: 2 (63.2/15.8)
##         :       installment_rate <= 2:
##         :       :...employment_length = > 7 yrs: 2 (9.4)
##         :           employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,
##         :           :                     unemployed}:
##         :           :...amount <= 1386: 2 (7.7/0.5)
##         :               amount > 1386: 1 (31.5/7.2)
##         purpose = radio/tv:
##         :...dependents > 1: 2 (8.5/1.6)
##         :   dependents <= 1:
##         :   :...employment_length = > 7 yrs: 1 (15.9/1.4)
##         :       employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,unemployed}:
##         :       :...housing = for free: 2 (4.2/0.5)
##         :           housing = rent: 1 (15.2/5.8)
##         :           housing = own:
##         :           :...months_loan_duration <= 39: 1 (68/30)
##         :               months_loan_duration > 39: 2 (7.4/0.5)
##         purpose = furniture:
##         :...installment_plan = stores: 1 (9.1)
##             installment_plan in {bank,none}:
##             :...amount > 4281: 2 (15.8/2.8)
##                 amount <= 4281:
##                 :...housing = for free: 1 (6.6/0.5)
##                     housing in {own,rent}:
##                     :...amount > 3573: 1 (17/3.4)
##                         amount <= 3573:
##                         :...personal_status = divorced male: 1 (7.5/2)
##                             personal_status in {married male,
##                             :                   single male}: 2 (25.6/10.2)
##                             personal_status = female:
##                             :...residence_history <= 1: 1 (4.1)
##                                 residence_history > 1:
##                                 :...age <= 37: 2 (30/6.1)
##                                     age > 37: 1 (4.1)
##
## -----  Trial 4:  -----
##
## Decision tree:
##
## months_loan_duration <= 7:
## :...amount <= 3380: 1 (48.6/5)
## :   amount > 3380: 2 (9.2/2.2)
## months_loan_duration > 7:
## :...savings_balance in {> 1000 DM,unknown}:
##     :...other_debtors = co-applicant: 1 (3.7)
##     :   other_debtors = guarantor: 2 (4.7/1.6)
##     :   other_debtors = none:
##     :   :...property in {building society savings,unknown/none}:
```

```
##     :       :...foreign_worker = no: 1 (2.5)
##     :       :   foreign_worker = yes:
##     :       :   :...savings_balance = > 1000 DM: 2 (15.8/3)
##     :       :       savings_balance = unknown:
##     :       :       :...installment_rate <= 1: 2 (7.2/1.2)
##     :       :           installment_rate > 1: 1 (42.5/12.1)
##     :       property in {other,real estate}:
##     :       :...savings_balance = > 1000 DM: 1 (19.3)
##     :           savings_balance = unknown:
##     :           :...residence_history > 3: 1 (25/1.6)
##     :               residence_history <= 3:
##     :               :...property = real estate: 2 (14.8/5.5)
##     :                   property = other:
##     :                   :...checking_balance = < 0 DM: 2 (6.4/1.2)
##     :                       checking_balance in {> 200 DM,1 - 200 DM,
##     :                                            unknown}: 1 (20.8/1.9)
##     savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}:
##     :...checking_balance in {> 200 DM,unknown}:
##         :...other_debtors = co-applicant: 2 (12.1/4.3)
##         :   other_debtors = guarantor: 1 (2.9)
##         :   other_debtors = none:
##         :   :...age > 48: 1 (17.2/1.2)
##         :       age <= 48:
##         :       :...purpose in {business,education,repairs}: 2 (36.9/15.9)
##         :           purpose in {car (used),domestic appliances,others,
##         :           :           retraining}: 1 (17.1/2.1)
##         :           purpose = car (new):
##         :           :...installment_plan in {bank,stores}: 2 (12.5/0.9)
##         :           :   installment_plan = none: 1 (21.1/6.4)
##         :           purpose = furniture:
##         :           :...months_loan_duration <= 30: 1 (31.8/8.5)
##         :           :   months_loan_duration > 30: 2 (7.7/0.9)
##         :           purpose = radio/tv:
##         :           :...months_loan_duration <= 9: 2 (8.7/0.4)
##         :               months_loan_duration > 9:
##         :               :...amount <= 2323: 1 (24.6)
##         :                   amount > 2323: [S1]
##         checking_balance in {< 0 DM,1 - 200 DM}:
##         :...months_loan_duration <= 22:
##             :...job = mangement self-employed: 1 (22.6/9.3)
##             :   job = unemployed non-resident: 2 (6.9/0.9)
##             :   job = unskilled resident:
##             :   :...age <= 54: 1 (58.5/14.7)
##             :   :   age > 54: 2 (7.5/0.9)
##             :   job = skilled employee:
##             :   :...credit_history = delayed: 1 (4.3/0.4)
##             :       credit_history = fully repaid this bank: 2 (4.8)
##             :       credit_history in {critical,fully repaid,repaid}:
##             :       :...amount <= 1381:
##             :           :...property in {other,unknown/none}: 2 (18.7/0.4)
##             :           :   property in {building society savings,real estate}:
##             :           :   :...foreign_worker = no: 1 (2)
##             :           :       foreign_worker = yes:
##             :           :       :...amount <= 662: 1 (5)
```

```
##             :           :           amount > 662: 2 (25.4/5.4)
##             :           amount > 1381:
##             :           :...employment_length in {4 - 7 yrs,
##             :           :                         unemployed}: 1 (13.3)
##             :           employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs}:
##             :           :...housing = for free: 2 (2.6)
##             :               housing = own: 1 (37.8/12.6)
##             :               housing = rent:
##             :               :...amount <= 1480: 1 (4)
##             :                   amount > 1480: 2 (22.5/4.4)
##             months_loan_duration > 22:
##             :...job = unemployed non-resident: 1 (1.4)
##                 job = unskilled resident: 2 (38.6/5.5)
##                 job in {mangement self-employed,skilled employee}:
##                 :...existing_credits > 1: 2 (63.2/17.9)
##                     existing_credits <= 1:
##                     :...personal_status in {divorced male,
##                         :                   married male}: 2 (17.1/4.4)
##                         personal_status = female:
##                         :...age <= 52: 2 (25.8/5)
##                         :   age > 52: 1 (2.2)
##                         personal_status = single male:
##                         :...other_debtors = co-applicant: 2 (4)
##                             other_debtors = guarantor: 1 (3.2)
##                             other_debtors = none:
##                             :...amount > 7596: 2 (14.2/3.1)
##                                 amount <= 7596:
##                                 :...installment_rate <= 2: 1 (11.6)
##                                     installment_rate > 2:
##                                     :...age <= 32: 1 (29.3/8.5)
##                                         age > 32: 2 (9.9/2.8)
##
## SubTree [S1]
##
## credit_history in {critical,fully repaid,fully repaid this bank}: 1 (6.7)
## credit_history in {delayed,repaid}:
## :...existing_credits <= 1: 1 (12.6/5.2)
##     existing_credits > 1: 2 (11/1.4)
##
## -----  Trial 5:  -----
##
## Decision tree:
##
## checking_balance = unknown:
## :...installment_plan = stores: 1 (14.6/5.4)
## :   installment_plan = bank:
## :   :...other_debtors in {co-applicant,guarantor}: 1 (3.1)
## :   :   other_debtors = none:
## :   :   :...existing_credits > 2: 1 (3.8)
## :   :       existing_credits <= 2:
## :   :       :...housing = for free: 1 (8.2/1.7)
## :   :           housing = rent: 2 (7/0.4)
## :   :           housing = own:
## :   :           :...telephone = yes: 2 (8.7/1.9)
```

```
## :   :               telephone = none:
## :   :               :...age <= 30: 1 (6)
## :   :                   age > 30: 2 (19.2/7)
## :   installment_plan = none:
## :   :...credit_history in {critical,fully repaid,
## :       :                  fully repaid this bank}: 1 (63.7/4)
## :       credit_history in {delayed,repaid}:
## :       :...existing_credits <= 1:
## :           :...purpose in {business,car (new),car (used),domestic appliances,
## :           :   :           education,others,radio/tv,retraining}: 1 (62.4/8.2)
## :           :   purpose in {furniture,repairs}: 2 (20/6.2)
## :           existing_credits > 1:
## :           :...employment_length = 4 - 7 yrs: 1 (7.6)
## :               employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,unemployed}:
## :               :...job in {mangement self-employed,
## :                   :       unemployed non-resident}: 2 (6.9)
## :                   job in {skilled employee,unskilled resident}:
## :                   :...employment_length in {> 7 yrs,0 - 1 yrs}: 2 (19.8/4.4)
## :                       employment_length in {1 - 4 yrs,
## :                                             unemployed}: 1 (7.2)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...property = unknown/none:
##     :...job = unskilled resident: 2 (10.7)
##     :   job in {mangement self-employed,skilled employee,
##     :   :       unemployed non-resident}:
##     :   :...installment_rate <= 2: 1 (31.5/11)
##     :       installment_rate > 2:
##     :       :...job = skilled employee: 2 (40.9/10.1)
##     :           job = unemployed non-resident: 1 (1)
##     :           job = mangement self-employed:
##     :           :...dependents > 1: 1 (2.2)
##     :               dependents <= 1:
##     :               :...residence_history <= 1: 1 (4.8/1)
##     :                   residence_history > 1: 2 (19.4/4.5)
##     property in {building society savings,other,real estate}:
##     :...purpose in {domestic appliances,others,repairs,
##         :           retraining}: 1 (28.8/11.1)
##         purpose = education: 2 (21.7/9.7)
##         purpose = car (used):
##         :...amount <= 7253: 1 (20.5/1)
##         :   amount > 7253: 2 (6.7/1.9)
##         purpose = business:
##         :...months_loan_duration <= 18: 1 (10.1)
##         :   months_loan_duration > 18:
##         :   :...housing = for free: 1 (0)
##         :       housing = rent: 2 (9.4/1.9)
##         :       housing = own:
##         :       :...savings_balance in {> 1000 DM,101 - 500 DM,501 - 1000 DM,
##         :           :                   unknown}: 1 (11.1)
##         :           savings_balance = < 100 DM:
##         :           :...amount <= 2292: 2 (7.7)
##         :               amount > 2292: 1 (17.4/7.2)
##         purpose = radio/tv:
##         :...months_loan_duration <= 8: 1 (6.8)
```

```
##         :   months_loan_duration > 8:
##         :   :...savings_balance = > 1000 DM: 2 (0)
##         :       savings_balance = unknown: 1 (15.1/2.5)
##         :       savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}:
##         :       :...months_loan_duration > 36: 2 (8.6)
##         :           months_loan_duration <= 36:
##         :           :...other_debtors = co-applicant: 2 (2.5/0.8)
##         :               other_debtors = guarantor: 1 (9.1/1.7)
##         :               other_debtors = none:
##         :               :...employment_length in {0 - 1 yrs,
##         :                   :                      unemployed}: 2 (25.9/5.8)
##         :                   employment_length in {> 7 yrs,
##         :                   :                      4 - 7 yrs}: 1 (22.2/5.7)
##         :                   employment_length = 1 - 4 yrs:
##         :                   :...months_loan_duration <= 15: 1 (21.4/8.1)
##         :                       months_loan_duration > 15: 2 (23.7/5)
##         purpose = furniture:
##         :...installment_plan = stores: 1 (6.1)
##         :   installment_plan in {bank,none}:
##         :   :...other_debtors = guarantor: 1 (4.3)
##         :       other_debtors in {co-applicant,none}:
##         :       :...savings_balance = > 1000 DM: 1 (5.1)
##         :           savings_balance in {101 - 500 DM,
##         :           :                   501 - 1000 DM}: 2 (4.1)
##         :           savings_balance in {< 100 DM,unknown}:
##         :           :...telephone = yes: 1 (30.4/9.6)
##         :               telephone = none:
##         :               :...personal_status = divorced male: 1 (4.3)
##         :                   personal_status in {married male,
##         :                   :                   single male}: 2 (33.4/9.9)
##         :                   personal_status = female:
##         :                   :...installment_plan = bank: 2 (2.7)
##         :                       installment_plan = none:
##         :                       :...months_loan_duration <= 9: 2 (3.1)
##         :                           months_loan_duration > 9: 1 (26.5/8.1)
##         purpose = car (new):
##         :...other_debtors in {co-applicant,guarantor}: 2 (12.4/2.8)
##             other_debtors = none:
##             :...property = real estate:
##                 :...installment_plan in {bank,stores}: 2 (2.7)
##                 :   installment_plan = none:
##                 :   :...amount > 4380: 1 (6)
##                 :       amount <= 4380:
##                 :       :...personal_status in {divorced male,
##                 :           :                   female}: 2 (7.3/0.4)
##                 :           personal_status in {married male,
##                 :                              single male}: 1 (29.7/6.1)
##                 property in {building society savings,other}:
##                 :...checking_balance = > 200 DM: 1 (3.7)
##                     checking_balance in {< 0 DM,1 - 200 DM}:
##                     :...amount <= 1126: 2 (19.7/0.4)
##                         amount > 1126:
##                         :...installment_plan = stores: 2 (0)
##                             installment_plan = bank: 1 (3.2)
```

```
##                             installment_plan = none:
##                             :...dependents > 1: 1 (5.9/1.2)
##                                 dependents <= 1:
##                                 :...job in {mangement self-employed,
##                                     :       unemployed non-resident,
##                                     :       unskilled resident}: 2 (19/3)
##                                     job = skilled employee:
##                                     :...installment_rate <= 1: 1 (4.9)
##                                         installment_rate > 1:
##                                         :...age <= 36: 2 (23.5/7.3)
##                                             age > 36: 1 (4.8)
##
## -----  Trial 6:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...foreign_worker = no: 1 (6.9)
## :   foreign_worker = yes:
## :   :...months_loan_duration <= 8: 1 (23.8/1.3)
## :       months_loan_duration > 8:
## :       :...job in {mangement self-employed,skilled employee,
## :           :       unemployed non-resident}:
## :           :...employment_length = > 7 yrs: 1 (67.6/8.6)
## :           :   employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,unemployed}:
## :           :   :...purpose in {car (used),domestic appliances,others,repairs,
## :           :       :           retraining}: 1 (21.8/2)
## :           :       purpose = education: 2 (16.3/8.1)
## :           :       purpose = business:
## :           :       :...existing_credits <= 2: 1 (23.5/8.6)
## :           :       :   existing_credits > 2: 2 (2.9)
## :           :       purpose = car (new):
## :           :       :...property in {building society savings,real estate,
## :           :       :   :            unknown/none}: 2 (20.1/5.9)
## :           :       :   property = other: 1 (4.1)
## :           :       purpose = furniture:
## :           :       :...months_loan_duration > 30: 2 (7.5/1.9)
## :           :       :   months_loan_duration <= 30:
## :           :       :   :...age <= 22: 2 (4.8/1.2)
## :           :       :       age > 22: 1 (18.5)
## :           :       purpose = radio/tv:
## :           :       :...dependents > 1: 1 (4.3)
## :           :           dependents <= 1:
## :           :           :...months_loan_duration <= 9: 2 (4.7)
## :           :               months_loan_duration > 9:
## :           :               :...installment_rate <= 1: 2 (2.1)
## :           :                   installment_rate > 1: 1 (38.2/9.1)
## :           :           job = unskilled resident:
## :           :           :...age > 48: 1 (6.3)
## :           :               age <= 48:
## :           :               :...purpose in {domestic appliances,others,
## :           :                   :           repairs}: 2 (0)
## :           :                   purpose in {business,retraining}: 1 (5.2)
## :           :                   purpose in {car (new),car (used),education,furniture,
```

```
## :                   :           radio/tv}:
## :                   :...installment_plan = bank: 2 (13.7/2.6)
## :                       installment_plan = stores: 1 (1.5)
## :                       installment_plan = none: [S1]
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...credit_history in {fully repaid,fully repaid this bank}:
##     :...other_debtors = co-applicant: 1 (3.3)
##     :   other_debtors in {guarantor,none}:
##     :   :...property in {building society savings,unknown/none}: 2 (36/3.1)
##     :       property in {other,real estate}:
##     :       :...housing in {for free,rent}: 2 (8/0.9)
##     :           housing = own:
##     :           :...age <= 35: 1 (23.4/8.2)
##     :               age > 35: 2 (7.1/0.8)
##     credit_history in {critical,delayed,repaid}:
##     :...other_debtors = guarantor: 1 (24.3/7.1)
##         other_debtors = co-applicant:
##         :...foreign_worker = no: 1 (3.5)
##         :   foreign_worker = yes:
##         :   :...installment_plan = stores: 2 (0)
##         :       installment_plan = bank: 1 (1.3)
##         :       installment_plan = none:
##         :       :...amount <= 1961: 1 (4.9)
##         :           amount > 1961: 2 (18.9/4.5)
##         other_debtors = none:
##         :...credit_history = delayed:
##             :...savings_balance in {101 - 500 DM,501 - 1000 DM,
##             :   :                   unknown}: 1 (22.9/2.7)
##             :   savings_balance in {< 100 DM,> 1000 DM}:
##             :   :...installment_rate <= 1: 1 (4.8)
##             :       installment_rate > 1:
##             :       :...job in {mangement self-employed,skilled employee,
##             :           :       unemployed non-resident}: 2 (21.6/1.9)
##             :           job = unskilled resident: 1 (3.5/0.8)
##             credit_history = critical:
##             :...residence_history <= 1: 1 (7.4)
##             :   residence_history > 1:
##             :   :...savings_balance in {> 1000 DM,101 - 500 DM,
##             :       :                   unknown}: 1 (16.4/2.2)
##             :       savings_balance = 501 - 1000 DM: 2 (5.1/2.2)
##             :       savings_balance = < 100 DM:
##             :       :...months_loan_duration > 36: 2 (6.3)
##             :           months_loan_duration <= 36:
##             :           :...personal_status in {divorced male,
##             :               :                   married male}: 2 (13.5/4.5)
##             :               personal_status in {female,
##             :                                   single male}: 1 (54.8/18.5)
##             credit_history = repaid:
##             :...savings_balance = > 1000 DM: 1 (6.2)
##                 savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,
##                 :                   unknown}:
##                 :...amount > 8086: 2 (22.1/1.8)
##                     amount <= 8086:
##                     :...purpose in {business,domestic appliances,
```

```
##                          :           retraining}: 2 (16.6/5)
##                          purpose in {car (used),education,others,
##                          :           repairs}: 1 (43.7/12.1)
##                          purpose = car (new):
##                          :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##                          :   :                      4 - 7 yrs}: 2 (56.2/20.9)
##                          :   employment_length = unemployed: 1 (5.7)
##                          purpose = furniture:
##                          :...residence_history <= 1: 1 (9.3/2.1)
##                          :   residence_history > 1:
##                          :   :...telephone = yes: 2 (16.5/6.8)
##                          :       telephone = none:
##                          :       :...months_loan_duration > 27: 2 (5.6)
##                          :           months_loan_duration <= 27:
##                          :           :...amount <= 2520: 2 (20.1/6.9)
##                          :               amount > 2520: 1 (11.4/1.6)
##                          purpose = radio/tv:
##                          :...amount > 5324: 2 (6.9)
##                              amount <= 5324:
##                              :...amount > 3190: 1 (9.8/0.3)
##                                  amount <= 3190: [S2]
##
## SubTree [S1]
##
## credit_history = fully repaid this bank: 2 (0)
## credit_history in {critical,fully repaid}: 1 (3.1)
## credit_history in {delayed,repaid}:
## :...amount <= 3229: 2 (25.1/4.1)
##     amount > 3229: 1 (3.5)
##
## SubTree [S2]
##
## property in {building society savings,unknown/none}: 2 (8.1/1.1)
## property = other:
## :...dependents <= 1: 1 (20.1/7.6)
## :   dependents > 1: 2 (4.1/0.8)
## property = real estate:
## :...months_loan_duration <= 11: 1 (4.7)
##     months_loan_duration > 11: 2 (20.4/4.3)
##
## -----  Trial 7:  -----
##
## Decision tree:
##
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...credit_history in {fully repaid,fully repaid this bank}:
## :   :...other_debtors = co-applicant: 1 (2.7)
## :   :   other_debtors in {guarantor,none}:
## :   :   :...age <= 22: 1 (3.8)
## :   :       age > 22: 2 (66.8/16.7)
## :   credit_history in {critical,delayed,repaid}:
## :   :...purpose in {car (used),others}: 1 (47.7/16.6)
## :       purpose in {domestic appliances,repairs,retraining}: 2 (26.3/10.1)
## :       purpose = business:
```

```
## :       :...personal_status = divorced male: 2 (4.4/0.6)
## :       :   personal_status in {female,married male,single male}: 1 (34.1/7.1)
## :       purpose = education:
## :       :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
## :       :   :                     unemployed}: 2 (25.4/5.2)
## :       :   employment_length = 4 - 7 yrs: 1 (5.4)
## :       purpose = furniture:
## :       :...dependents > 1: 1 (6.1/0.5)
## :       :   dependents <= 1:
## :       :   :...savings_balance in {> 1000 DM,unknown}: 1 (21.7/7.5)
## :       :       savings_balance in {101 - 500 DM,501 - 1000 DM}: 2 (6.6/1.5)
## :       :       savings_balance = < 100 DM:
## :       :       :...personal_status = married male: 1 (5.1)
## :       :           personal_status in {divorced male,female,single male}:
## :       :           :...amount <= 1893: 1 (25.1/5)
## :       :               amount > 1893: 2 (54.1/17.9)
## :       purpose = car (new):
## :       :...installment_plan in {bank,stores}: 2 (19.7/4.3)
## :       :   installment_plan = none:
## :       :   :...job = mangement self-employed: 2 (15.8/5.9)
## :       :       job in {skilled employee,unemployed non-resident,
## :       :       :       unskilled resident}:
## :       :       :...checking_balance = 1 - 200 DM: 1 (40.4/8.8)
## :       :           checking_balance = < 0 DM:
## :       :           :...installment_rate <= 2: 1 (17.7/3.3)
## :       :               installment_rate > 2:
## :       :               :...telephone = none: 2 (30.3/8)
## :       :                   telephone = yes: 1 (10.1/2.1)
## :       purpose = radio/tv:
## :       :...foreign_worker = no: 1 (3.1)
## :           foreign_worker = yes:
## :           :...months_loan_duration <= 8: 1 (6.8)
## :               months_loan_duration > 8:
## :               :...employment_length = > 7 yrs: 1 (15/4.1)
## :                   employment_length in {4 - 7 yrs,unemployed}: 2 (20.6/7)
## :                   employment_length = 1 - 4 yrs:
## :                   :...credit_history in {critical,repaid}: 2 (33.8/13.6)
## :                   :   credit_history = delayed: 1 (3.3)
## :                   employment_length = 0 - 1 yrs:
## :                   :...other_debtors = co-applicant: 2 (0)
## :                       other_debtors = guarantor: 1 (1.6)
## :                       other_debtors = none:
## :                       :...amount <= 2214: 2 (14.4)
## :                           amount > 2214: 1 (12.4/4.6)
## checking_balance in {> 200 DM,unknown}:
## :...foreign_worker = no: 1 (5.6)
##     foreign_worker = yes:
##     :...installment_plan = stores: 2 (17.4/7.6)
##         installment_plan = bank:
##         :...housing in {for free,own}: 1 (55/21.3)
##         :   housing = rent: 2 (5.4)
##         installment_plan = none:
##         :...credit_history in {critical,fully repaid,
##             :                  fully repaid this bank}: 1 (69.3/11.6)
```

```
##             credit_history = delayed:
##             :...residence_history <= 1: 2 (3.5)
##             :   residence_history > 1:
##             :   :...installment_rate <= 3: 1 (9.2)
##             :       installment_rate > 3: 2 (21.3/7.6)
##             credit_history = repaid:
##             :...telephone = yes: 1 (49.7/6.8)
##                 telephone = none:
##                 :...other_debtors in {co-applicant,guarantor}: 2 (11.3/3.3)
##                     other_debtors = none:
##                     :...savings_balance in {> 1000 DM,unknown}: 1 (11.2)
##                         savings_balance in {< 100 DM,101 - 500 DM,
##                         :                   501 - 1000 DM}:
##                         :...personal_status in {divorced male,
##                             :                   married male}: 1 (7.8)
##                             personal_status in {female,single male}:
##                             :...housing = for free: 2 (2.2/0.5)
##                                 housing = rent: 1 (10/2.5)
##                                 housing = own:
##                                 :...age <= 34: 2 (32.8/12.5)
##                                     age > 34: 1 (8)
##
## -----  Trial 8:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...installment_plan = bank:
## :   :...other_debtors = guarantor: 2 (0)
## :   :   other_debtors = co-applicant: 1 (1.7)
## :   :   other_debtors = none:
## :   :   :...existing_credits > 2: 1 (3.1)
## :   :       existing_credits <= 2:
## :   :       :...savings_balance in {< 100 DM,501 - 1000 DM,
## :   :           :                   unknown}: 2 (47.7/16.8)
## :   :           savings_balance in {> 1000 DM,101 - 500 DM}: 1 (9/1.6)
## :   installment_plan in {none,stores}:
## :   :...purpose in {car (used),domestic appliances,education,others,
## :       :           retraining}: 1 (39.1/4.1)
## :       purpose = repairs: 2 (7.8/3.5)
## :       purpose = business:
## :       :...job = mangement self-employed: 2 (7.9/0.7)
## :       :   job in {skilled employee,unemployed non-resident,
## :       :           unskilled resident}: 1 (18.7/4.2)
## :       purpose = car (new):
## :       :...existing_credits <= 2: 1 (50/7.7)
## :       :   existing_credits > 2: 2 (3.4/0.6)
## :       purpose = furniture:
## :       :...job in {mangement self-employed,
## :       :   :       unemployed non-resident}: 2 (5.7/1.9)
## :       :   job in {skilled employee,unskilled resident}: 1 (49.3/11.7)
## :       purpose = radio/tv:
## :       :...checking_balance = > 200 DM:
## :           :...age <= 41: 2 (19.4/5.9)
```

```
## :           :   age > 41: 1 (4.8)
## :           checking_balance = unknown:
## :           :...age <= 23: 2 (6.6/1.7)
## :               age > 23: 1 (38.6/4.2)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...employment_length = unemployed:
##     :...residence_history <= 1: 2 (5.5)
##     :   residence_history > 1:
##     :   :...dependents <= 1: 1 (39.3/9.7)
##     :       dependents > 1: 2 (6.6/1.5)
##     employment_length = 4 - 7 yrs:
##     :...age > 29: 1 (61.5/13.3)
##     :   age <= 29:
##     :   :...installment_rate <= 1: 1 (3.6)
##     :       installment_rate > 1:
##     :       :...savings_balance in {< 100 DM,> 1000 DM,101 - 500 DM,
##     :           :                   501 - 1000 DM}: 2 (32.7/8.8)
##     :           savings_balance = unknown: 1 (2.5)
##     employment_length = 0 - 1 yrs:
##     :...foreign_worker = no: 1 (5.5)
##     :   foreign_worker = yes:
##     :   :...housing = for free: 1 (7.5/2.5)
##     :       housing = rent: 2 (32.9/7.3)
##     :       housing = own:
##     :       :...savings_balance in {> 1000 DM,501 - 1000 DM,
##     :           :                   unknown}: 1 (7.9)
##     :           savings_balance in {< 100 DM,101 - 500 DM}:
##     :           :...residence_history <= 1: 1 (29/9.7)
##     :               residence_history > 1: 2 (33.5/8.4)
##     employment_length = 1 - 4 yrs:
##     :...amount > 7721: 2 (13.6/0.6)
##     :   amount <= 7721:
##     :   :...housing = for free: 2 (6.7/2.9)
##     :       housing = rent:
##     :       :...residence_history <= 3: 1 (10.3/4)
##     :       :   residence_history > 3: 2 (26/7.9)
##     :       housing = own:
##     :       :...personal_status = divorced male: 1 (10.7/1.6)
##     :           personal_status = married male:
##     :           :...job = skilled employee: 2 (16.5/6.7)
##     :           :   job in {mangement self-employed,unemployed non-resident,
##     :           :           unskilled resident}: 1 (7.3)
##     :           personal_status = single male:
##     :           :...amount <= 902: 2 (7.5/1.4)
##     :           :   amount > 902: 1 (59.1/13.3)
##     :           personal_status = female:
##     :           :...residence_history <= 1: 1 (7.4/0.9)
##     :               residence_history > 1:
##     :               :...age <= 37: 2 (29.9/8.7)
##     :                   age > 37: 1 (5.4)
##     employment_length = > 7 yrs:
##     :...personal_status = married male: 1 (4.8)
##         personal_status in {divorced male,female,single male}:
##         :...months_loan_duration > 40: 2 (6)
```

```
##             months_loan_duration <= 40:
##             :...residence_history <= 3:
##                 :...savings_balance in {< 100 DM,> 1000 DM,501 - 1000 DM,
##                 :   :                  unknown}: 2 (27.3/3.9)
##                 :   savings_balance = 101 - 500 DM: 1 (3.9/0.5)
##                 residence_history > 3:
##                 :...age <= 30: 1 (13.7/0.6)
##                     age > 30:
##                     :...existing_credits <= 1: 2 (36.3/9.5)
##                         existing_credits > 1: [S1]
##
## SubTree [S1]
##
## credit_history in {critical,fully repaid this bank,repaid}: 1 (20.9/4.5)
## credit_history in {delayed,fully repaid}: 2 (3.9)
##
## -----  Trial 9:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...checking_balance = > 200 DM:
## :   :...dependents <= 1: 1 (60.2/17.5)
## :   :   dependents > 1: 2 (9.4/2.7)
## :   checking_balance = unknown:
## :   :...amount <= 4455: 1 (163.6/30.7)
## :       amount > 4455:
## :       :...employment_length in {> 7 yrs,4 - 7 yrs}: 1 (20.2)
## :           employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}: 2 (44.6/13.8)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...foreign_worker = no: 1 (14.6/3.4)
##     foreign_worker = yes:
##     :...credit_history in {fully repaid,fully repaid this bank}: 2 (71.9/23.9)
##         credit_history in {critical,delayed,repaid}:
##         :...amount > 7966:
##             :...credit_history in {critical,repaid}: 2 (31.9/5.2)
##             :   credit_history = delayed: 1 (4.4/1.4)
##             amount <= 7966:
##             :...installment_plan = stores: 2 (20.7/6.4)
##                 installment_plan in {bank,none}:
##                 :...months_loan_duration > 36:
##                     :...dependents > 1: 1 (6.3/1.6)
##                     :   dependents <= 1:
##                     :   :...employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,
##                     :       :                     4 - 7 yrs}: 2 (24/2.3)
##                     :       employment_length = unemployed: 1 (3.4)
##                     months_loan_duration <= 36:
##                     :...other_debtors = co-applicant: 2 (17.9/8.4)
##                         other_debtors = guarantor: 1 (22.1/4.4)
##                         other_debtors = none:
##                         :...employment_length = 4 - 7 yrs:
##                             :...personal_status in {divorced male,
##                             :   :                   married male}: 2 (13.8/5)
##                             :   personal_status in {female,
```

```
##                                :                        single male}: 1 (41.6/4.7)
##                               employment_length = unemployed:
##                               :...residence_history <= 2: 2 (14.9/2.1)
##                               :   residence_history > 2: 1 (19.1/4.6)
##                               employment_length = 1 - 4 yrs:
##                               :...housing in {for free,own}: 1 (95.8/31.1)
##                               :   housing = rent:
##                               :   :...purpose in {car (new),
##                               :       :           car (used)}: 1 (14.8/3.2)
##                               :       purpose in {business,domestic appliances,
##                               :                   education,furniture,others,
##                               :                   radio/tv,repairs,
##                               :                   retraining}: 2 (13.6/1.2)
##                               employment_length = > 7 yrs:
##                               :...months_loan_duration <= 8: 1 (7.3)
##                               :   months_loan_duration > 8:
##                               :   :...residence_history <= 3:
##                               :       :...amount <= 5129: 2 (21.1/4.9)
##                               :       :   amount > 5129: 1 (3.3)
##                               :       residence_history > 3:
##                               :       :...amount <= 6948: 1 (46.9/14.4)
##                               :           amount > 6948: 2 (3.9/0.9)
##                               employment_length = 0 - 1 yrs:
##                               :...job in {mangement self-employed,
##                                   :       unemployed non-resident}: 1 (7.9/2.2)
##                                   job = unskilled resident: 2 (21.3/7.4)
##                                   job = skilled employee:
##                                   :...amount > 4870: 1 (6.5)
##                                       amount <= 4870:
##                                       :...existing_credits > 1: 2 (4.6/0.5)
##                                           existing_credits <= 1: [S1]
##
## SubTree [S1]
##
## personal_status in {divorced male,single male}: 1 (10.5)
## personal_status in {female,married male}:
## :...credit_history = delayed: 2 (0)
##     credit_history = critical: 1 (1.8)
##     credit_history = repaid:
##     :...months_loan_duration <= 24: 2 (25.9/8.1)
##         months_loan_duration > 24: 1 (3.1)
##
##
## Evaluation on training data (900 cases):
##
## Trial        Decision Tree
## -----      ----------------
##    Size      Errors
##
##    0     54  135(15.0%)
##    1     37  184(20.4%)
##    2     58  172(19.1%)
##    3     40  173(19.2%)
##    4     54  188(20.9%)
```

```
##    5     63  162(18.0%)
##    6     61  158(17.6%)
##    7     46  209(23.2%)
##    8     49  186(20.7%)
##    9     35  178(19.8%)
## boost             29( 3.2%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     630     3    (a): class 1
##      26   241    (b): class 2
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##  100.00% months_loan_duration
##  100.00% foreign_worker
##   99.00% employment_length
##   98.67% purpose
##   98.00% other_debtors
##   96.67% amount
##   96.44% savings_balance
##   95.22% installment_plan
##   93.67% credit_history
##   90.00% job
##   87.11% installment_rate
##   74.44% age
##   74.33% property
##   59.33% existing_credits
##   58.56% residence_history
##   55.33% personal_status
##   54.89% housing
##   46.00% dependents
##   37.44% telephone
##
##
## Time: 0.1 secs
```

```
# Making prediction using the new model
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 |         1 |         2 | Row Total |
## ---------------|-----------|-----------|-----------|
##              1 |        60 |         7 |        67 |
##                |     0.600 |     0.070 |           |
## ---------------|-----------|-----------|-----------|
##              2 |        17 |        16 |        33 |
##                |     0.170 |     0.160 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        77 |        23 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

```
# Defining cost matirx dimensions
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"
```

```
# Assigning penalty costs
error_cost <- matrix(c(0, 1, 4, 0), nrow = 2)
error_cost
```

```
##      [,1] [,2]
## [1,]    0    4
## [2,]    1    0
```

```
# Applying decision tree using cost parameter and making predictions
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,
 dnn = c('actual default', 'predicted default'))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  100
```

```
##
##
##                | predicted default
## actual default |         1 |         2 | Row Total |
## ---------------|-----------|-----------|-----------|
##              1 |        33 |        34 |        67 |
##                |     0.330 |     0.340 |           |
## ---------------|-----------|-----------|-----------|
##              2 |         7 |        26 |        33 |
##                |     0.070 |     0.260 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        40 |        60 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

Quesiton 2

Step 1 – Collecting data

Downloading the mushroom dataset

Step 2 - Exploring and preparing the data

Exploring the dataset and setting the veil_type to null since this dataset contains only one type of that variable (1 level only)

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

```
mushrooms$veil_type <- NULL
# Displaying the distribution of the two types of mushrooms
table(mushrooms$type)
```

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

Step 3 – Training a model on the data

Using the OneR function to train the model using type as the categorical variable

```
Sys.setenv(JAVA_HOME='C:\\Program Files\\Java\\jre1.8.0_201')
library(rJava)
library(RWeka)
mushroom_1R <- OneR(type ~ ., data = mushrooms)
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)
```

Step 4 – Evaluating model performance

The model correctly predicts 8004 out of 8124 samples, which is ~98.5% of the samples.

```
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
```

Step 5 – Improving model performance

Using the JRip function to improve the model. What JRip does is it creates a set of 9 rules based on the predictors, which are used in making predictions. As can be seen in the summary, it gives 100% accuracy.

```
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
```

```
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
```

Question 3

k-NN Algorithm

In this algorithm, first the distances are calculated between the given values and the unknown value. These differences are then ordered and the mode of k (user defined) number of smallest distances is considered as the prediction for the unknown value.

- Uses - kNN can be used for computer vision applications, predicting whether a person will enjoy a book or for identifying patterns in genetic data.
- Strength - well-suited for classification tasks where a concept is difficult to define, yet the items of similar class type tend to be fairly homogeneous.

- Weakness - this algorithm will struggle to identify class boundaries if the data is noisy and the different groups cannot be easily distinguished.

Naive Bayes Classification

In this classification model, a likelihood table is created for the categorical class variable given independent predictor variables using the Bayes rule. It is called naive, because it makes a lot of naive assumptions.

- Uses - Naive Bayes can be used for text classification such as spam filtering, anomaly detection in computer networks or for diagnosing medical conditions given a set of observed symptoms.
- Strengths - it is simple, fast and effective, even with noisy or missing data, requires fewer examples for training and it is also easy to obtain estimated probablity for a prediction
- Weakness - relies on an often-faulty assumption of equally important and independent features, not ideal for datasets with many numeric features, estimated probabilities are less reliable than the predicted classes

C5.0 Decision Trees

Decision trees utilize a tree structure to model the relationships among the features and the potential outcomes. The C5.0 uses the Quinlan's C5.0 algorithm to fit the classification model

- Uses - can be used for credit scorig models with a clear criteria to accept or reject an applicant, marketing studies of customer behavior or for diagnosis of medical conditions based on laboratory measurements, or the rate of disease progression
- Strengths - highly automatic learning process, excludes unimportant features, results in a model that can be interpreted without a mathematical background (for relatively small trees)
- Weakness - models are often biased toward splits on features having a large number of levels, easy to over- or under- fit the model, large trees can be difficult to interpret and the decisions they make may seem counterintuitive

RIPPER Rules

RIPPER stands for Repeated Incremental Pruning to Produce Error Reduction. What RIPPER does is it creates a set of rules based on the predictors, which are used in making predictions.

- Uses - Identifying conditions that lead to a mechanical failure in devices, to describe key characteristics of groups for segmentation or to find conditions preceding large fluctuations in stock market.
- Strengths - Generates easy-to-understand, human-readable rules , efficiently handles large and noisy datasets, model is usually simpler than comparative decision trees
- Weakness - rules generated might defy common sense or expert knowledge, not ideal while working with numeric data, performance might not be as good as more complex models

Question 4

Model Ensembles

A model ensembles are based on the idea that by combining multiple weaker learning models, a stronger learning model is created. To create such an ensemble, first the input training data is used to build a number of models. An allocation function dictates how much of the training data each of the model receives. Then, these models are used to generate a set of predictions. A combination function governs how disagreements among the predictions are reconciled.

Importance and benefits - Model ensembles offer the following performance advantages over single models - * Future problems can be better generalized - incorporating multiple models' opinions, reduces the chances of overfitting. * Better performance for small as well as large datasets * Ability to synthesize data from distinct domains * A more nuanced understanding of difficult learning tasks

Bagging - Bootstrap aggregating or bagging generates a number of training datasets by bootstrap sampling the original training data.These datasets are then used to generate a set of models using a single learning algorithm. The models' predictions are combined using voting (for classification) or averaging (for numeric prediction). Bagging can perform quite well as long as it is used with relatively unstable learners. Unstable models are essential in order to ensure the ensemble's diversity in spite of only minor variations between the bootstrap training datasets. For this reason, bagging is often used with decision trees, which have the tendency to vary dramatically given minor changes in the input data.

Boosting - Boosting is a ensemble-based method which boosts the performance of weak learners to attain the performance of stronger learners. Boosting uses ensembles of models trained on resampled data and a vote to determine the final prediction. There are two key distinctions. First, the resampled datasets in boosting are constructed specifically to generate complementary learners. Second, rather than giving each learner an equal vote, boosting gives each learner's vote a weight based on its past performance. Models that perform better have greater influence over the ensemble's final prediction. Since the models in the ensemble are built to be complementary, it is possible to increase ensemble performance to an arbitrary threshold simply by adding additional classifiers to the group, assuming that each classifier performs better than random chance.

Reference - Machine Learning with R