Practice 5-DA5030

Shivam Verma

```
library(C50)
library(gmodels)
```

Problem 1.

Step 2 Exploring and Preparing the data

```
credit <- read.csv("credit.csv")
credit$default <- ifelse(credit$default==2, "Yes", "No")
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             : chr  "No" "Yes" "No" "No" ...
##  $ 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)
```

```
##
##  No Yes
## 700 300
```

Data preparation - creating random training and test datasets

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

```
##  int [1:900] 415 463 179 526 195 938 818 118 299 229 ...
```

```
# creating train and test datasets
credit_train <- credit[train_sample,]
credit_test <- credit[-train_sample,]
label <- as.factor(credit_train$default)
prop.table(table(credit_train$default))
```

```
##
##        No       Yes
## 0.7055556 0.2944444
```

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

```
##
##   No  Yes
## 0.65 0.35
```

Step 3 - training a model on the data

```
# passing credit_train without default column
credit_model <- C5.0(credit_train[-17], label)
credit_model
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = label)
##
## Classification Tree
## Number of samples: 900
## Number of predictors: 20
##
## Tree size: 42
##
## Non-standard options: attempt to group attributes
```

```
summary(credit_model)
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = label)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Jun 23 11:45:14 2020
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 900 cases (21 attributes) from undefined.data
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}: No (412/54)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...credit_history in {fully repaid,fully repaid this bank}:
##     :...housing = rent: Yes (16/1)
##     :   housing = for free:
##     :   :...other_debtors = co-applicant: No (2)
##     :   :   other_debtors in {guarantor,none}: Yes (12/1)
##     :   housing = own:
##     :   :...purpose in {business,car (used),retraining}: No (10/2)
##     :       purpose in {domestic appliances,education,others,radio/tv,
##     :       :           repairs}: Yes (6/1)
##     :       purpose = car (new):
##     :       :...months_loan_duration <= 22: Yes (6)
##     :       :   months_loan_duration > 22: No (2)
##     :       purpose = furniture:
##     :       :...installment_plan = bank: Yes (5/1)
##     :           installment_plan in {none,stores}: No (4)
##     credit_history in {critical,delayed,repaid}:
##     :...months_loan_duration <= 15: No (180/45)
##         months_loan_duration > 15:
##         :...savings_balance in {> 1000 DM,unknown}:
##             :...credit_history in {critical,delayed}: No (14)
##             :   credit_history = repaid:
##             :   :...purpose in {business,domestic appliances,education,others,
##             :       :           repairs,retraining}: No (5)
##             :       purpose = car (new): Yes (7/1)
```

```
##             :       purpose = car (used):
##             :       :...amount <= 6967: No (4)
##             :       :   amount > 6967: Yes (2)
##             :       purpose = furniture:
##             :       :...age <= 27: Yes (2)
##             :       :   age > 27: No (5)
##             :       purpose = radio/tv:
##             :       :...amount <= 6110: No (5)
##             :           amount > 6110: Yes (2)
##             savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}:
##             :...months_loan_duration > 47: Yes (23/3)
##                 months_loan_duration <= 47:
##                 :...employment_length = 0 - 1 yrs:
##                     :...residence_history <= 1: No (16/6)
##                     :   residence_history > 1: Yes (27/6)
##                     employment_length = unemployed:
##                     :...residence_history <= 2: Yes (7)
##                     :   residence_history > 2: No (12/2)
##                     employment_length = > 7 yrs:
##                     :...purpose = car (new): Yes (11/3)
##                     :   purpose in {car (used),domestic appliances,education,
##                     :   :           others,radio/tv,repairs,
##                     :   :           retraining}: No (13/1)
##                     :   purpose = business:
##                     :   :...personal_status in {divorced male,female}: Yes (3)
##                     :   :   personal_status in {married male,
##                     :   :                       single male}: No (3)
##                     :   purpose = furniture:
##                     :   :...job = mangement self-employed: Yes (2)
##                     :       job in {skilled employee,unemployed non-resident,
##                     :               unskilled resident}: No (5/1)
##                     employment_length = 1 - 4 yrs:
##                     :...installment_rate > 3: Yes (20/3)
##                     :   installment_rate <= 3:
##                     :   :...other_debtors = co-applicant: Yes (3)
##                     :       other_debtors = guarantor: No (2)
##                     :       other_debtors = none:
##                     :       :...checking_balance = 1 - 200 DM: No (8/1)
##                     :           checking_balance = < 0 DM: [S1]
##                     employment_length = 4 - 7 yrs:
##                     :...savings_balance in {101 - 500 DM,
##                         :                   501 - 1000 DM}: No (8)
##                         savings_balance = < 100 DM:
##                         :...job in {mangement self-employed,
##                             :       unemployed non-resident,
##                             :       unskilled resident}: No (6)
##                             job = skilled employee:
##                             :...dependents > 1: No (3/1)
##                                 dependents <= 1:
##                                 :...months_loan_duration <= 22: No (3)
##                                     months_loan_duration > 22: Yes (8)
##
## SubTree [S1]
##
```

```
## personal_status in {divorced male,married male}: No (3)
## personal_status in {female,single male}: Yes (13/3)
##
##
## Evaluation on training data (900 cases):
##
##      Decision Tree
##    ----------------
##    Size      Errors
##
##      42  136(15.1%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     612    23    (a): class No
##     113   152    (b): class Yes
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##   54.22% credit_history
##   48.11% months_loan_duration
##   27.22% savings_balance
##   19.56% employment_length
##   11.33% purpose
##    7.00% housing
##    6.89% residence_history
##    5.44% installment_rate
##    4.78% other_debtors
##    3.00% job
##    2.44% personal_status
##    1.56% dependents
##    1.44% amount
##    1.00% installment_plan
##    0.78% age
##
##
## Time: 0.0 secs
```

Step 4 - evaluating model performance

```
credit_pred <- predict(credit_model, credit_test)
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 |        55 |        10 |        65 |
##                |     0.550 |     0.100 |           |
## ---------------|-----------|-----------|-----------|
##            Yes |        22 |        13 |        35 |
##                |     0.220 |     0.130 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        77 |        23 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

Accuracy is 68%

Step 5 - Improving model performance

```
# applying adaptive boosting
credit_boost10 <- C5.0(credit_train[-17], label, trials = 10)
credit_boost10
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = label, trials = 10)
##
## Classification Tree
## Number of samples: 900
## Number of predictors: 20
##
## Number of boosting iterations: 10
## Average tree size: 36.3
##
## Non-standard options: attempt to group attributes
```

```
summary(credit_boost10)
```

```
##
## Call:
## C5.0.default(x = credit_train[-17], y = label, trials = 10)
##
##
## C5.0 [Release 2.07 GPL Edition]      Tue Jun 23 11:45:14 2020
## -------------------------------
##
## Class specified by attribute `outcome'
##
## Read 900 cases (21 attributes) from undefined.data
```

```
##
## -----  Trial 0:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}: No (412/54)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...credit_history in {fully repaid,fully repaid this bank}:
##     :...housing = rent: Yes (16/1)
##     :   housing = for free:
##     :   :...other_debtors = co-applicant: No (2)
##     :   :   other_debtors in {guarantor,none}: Yes (12/1)
##     :   housing = own:
##     :   :...purpose in {business,car (used),retraining}: No (10/2)
##     :       purpose in {domestic appliances,education,others,radio/tv,
##     :       :           repairs}: Yes (6/1)
##     :       purpose = car (new):
##     :       :...months_loan_duration <= 22: Yes (6)
##     :       :   months_loan_duration > 22: No (2)
##     :       purpose = furniture:
##     :       :...installment_plan = bank: Yes (5/1)
##     :           installment_plan in {none,stores}: No (4)
##     credit_history in {critical,delayed,repaid}:
##     :...months_loan_duration <= 15: No (180/45)
##         months_loan_duration > 15:
##         :...savings_balance in {> 1000 DM,unknown}:
##             :...credit_history in {critical,delayed}: No (14)
##             :   credit_history = repaid:
##             :   :...purpose in {business,domestic appliances,education,others,
##             :       :           repairs,retraining}: No (5)
##             :       purpose = car (new): Yes (7/1)
##             :       purpose = car (used):
##             :       :...amount <= 6967: No (4)
##             :       :   amount > 6967: Yes (2)
##             :       purpose = furniture:
##             :       :...age <= 27: Yes (2)
##             :       :   age > 27: No (5)
##             :       purpose = radio/tv:
##             :       :...amount <= 6110: No (5)
##             :           amount > 6110: Yes (2)
##             savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}:
##             :...months_loan_duration > 47: Yes (23/3)
##                 months_loan_duration <= 47:
##                 :...employment_length = 0 - 1 yrs:
##                     :...residence_history <= 1: No (16/6)
##                     :   residence_history > 1: Yes (27/6)
##                     employment_length = unemployed:
##                     :...residence_history <= 2: Yes (7)
##                     :   residence_history > 2: No (12/2)
##                     employment_length = > 7 yrs:
##                     :...purpose = car (new): Yes (11/3)
##                     :   purpose in {car (used),domestic appliances,education,
##                     :   :           others,radio/tv,repairs,
##                     :   :           retraining}: No (13/1)
```

```
##                     :   purpose = business:
##                     :   :...personal_status in {divorced male,female}: Yes (3)
##                     :   :   personal_status in {married male,
##                     :   :                       single male}: No (3)
##                     :   purpose = furniture:
##                     :   :...job = mangement self-employed: Yes (2)
##                     :       job in {skilled employee,unemployed non-resident,
##                     :               unskilled resident}: No (5/1)
##                     employment_length = 1 - 4 yrs:
##                     :...installment_rate > 3: Yes (20/3)
##                     :   installment_rate <= 3:
##                     :   :...other_debtors = co-applicant: Yes (3)
##                     :       other_debtors = guarantor: No (2)
##                     :       other_debtors = none:
##                     :       :...checking_balance = 1 - 200 DM: No (8/1)
##                     :           checking_balance = < 0 DM: [S1]
##                     employment_length = 4 - 7 yrs:
##                     :...savings_balance in {101 - 500 DM,
##                         :                   501 - 1000 DM}: No (8)
##                         savings_balance = < 100 DM:
##                         :...job in {mangement self-employed,
##                             :       unemployed non-resident,
##                             :       unskilled resident}: No (6)
##                             job = skilled employee:
##                             :...dependents > 1: No (3/1)
##                                 dependents <= 1:
##                                 :...months_loan_duration <= 22: No (3)
##                                     months_loan_duration > 22: Yes (8)
##
## SubTree [S1]
##
## personal_status in {divorced male,married male}: No (3)
## personal_status in {female,single male}: Yes (13/3)
##
## -----  Trial 1:  -----
##
## Decision tree:
##
## checking_balance = unknown:
## :...installment_plan = none: No (264.1/49.6)
## :   installment_plan in {bank,stores}:
## :   :...other_debtors in {co-applicant,guarantor}: No (3.2)
## :       other_debtors = none:
## :       :...employment_length in {> 7 yrs,4 - 7 yrs}: No (29.3/8.6)
## :           employment_length in {0 - 1 yrs,1 - 4 yrs,
## :                                 unemployed}: Yes (40.5/10.3)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...other_debtors = guarantor: No (35.3/7.2)
##     other_debtors in {co-applicant,none}:
##     :...savings_balance in {> 1000 DM,501 - 1000 DM,unknown}:
##         :...amount > 1530: No (63.4/13.8)
##         :   amount <= 1530:
##         :   :...installment_rate <= 2: No (6.4)
##         :       installment_rate > 2:
```

```
##         :       :...dependents > 1: Yes (5.1)
##         :           dependents <= 1:
##         :           :...months_loan_duration <= 11: No (5.6)
##         :               months_loan_duration > 11: Yes (25.6/7.9)
##         savings_balance in {< 100 DM,101 - 500 DM}:
##         :...credit_history in {critical,delayed}:
##             :...savings_balance = 101 - 500 DM: No (16.2/2.4)
##             :   savings_balance = < 100 DM:
##             :   :...other_debtors = co-applicant: Yes (7.5/2.4)
##             :       other_debtors = none:
##             :       :...personal_status in {divorced male,
##             :           :                   married male}: Yes (13.8/4)
##             :           personal_status = female: No (26.9/10.7)
##             :           personal_status = single male:
##             :           :...installment_rate <= 1: No (9.3)
##             :               installment_rate > 1:
##             :               :...credit_history = critical: No (38.8/10.7)
##             :                   credit_history = delayed: Yes (14.4/3.2)
##             credit_history in {fully repaid,fully repaid this bank,repaid}:
##             :...amount > 11054: Yes (16.9/0.8)
##                 amount <= 11054:
##                 :...job = mangement self-employed: No (36.5/13.6)
##                     job = unemployed non-resident: Yes (4.5)
##                     job in {skilled employee,unskilled resident}:
##                     :...installment_rate <= 2:
##                         :...dependents > 1: Yes (11.2/2.4)
##                         :   dependents <= 1:
##                         :   :...installment_rate <= 1: Yes (24.8/9.3)
##                         :       installment_rate > 1: No (42.6/14.4)
##                         installment_rate > 2:
##                         :...personal_status in {divorced male,female,
##                             :                   married male}: Yes (79.5/19.9)
##                             personal_status = single male:
##                             :...savings_balance = 101 - 500 DM: No (9.1/1.6)
##                                 savings_balance = < 100 DM:
##                                 :...months_loan_duration <= 11: No (9.9/2.2)
##                                     months_loan_duration > 11: Yes (59.6/13.8)
##
## -----  Trial 2:  -----
##
## Decision tree:
##
## foreign_worker = no: No (27.8/3.9)
## foreign_worker = yes:
## :...checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
##     :...property = unknown/none:
##     :   :...housing in {own,rent}: Yes (31.8/5.2)
##     :   :   housing = for free:
##     :   :   :...dependents > 1: Yes (23.5/5.4)
##     :   :       dependents <= 1:
##     :   :       :...employment_length in {0 - 1 yrs,4 - 7 yrs,
##     :   :           :                     unemployed}: No (18.3/2.4)
##     :   :           employment_length in {> 7 yrs,1 - 4 yrs}:
##     :   :           :...savings_balance in {< 100 DM,> 1000 DM,501 - 1000 DM,
```

```
##     :   :               :                   unknown}: Yes (31.3/7.9)
##     :   :               savings_balance = 101 - 500 DM: No (4.5/0.7)
##     :   property in {building society savings,other,real estate}:
##     :   :...purpose in {business,car (used),others,repairs,
##     :       :           retraining}: No (81.6/25.7)
##     :       purpose in {domestic appliances,education}: Yes (28.2/10.3)
##     :       purpose = radio/tv:
##     :       :...months_loan_duration > 36: Yes (15.1/1.3)
##     :       :   months_loan_duration <= 36:
##     :       :   :...credit_history in {critical,delayed,fully repaid this bank,
##     :       :       :                  repaid}: No (112.3/35.6)
##     :       :       credit_history = fully repaid: Yes (4.1)
##     :       purpose = car (new):
##     :       :...savings_balance = > 1000 DM: No (4.8)
##     :       :   savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM,
##     :       :   :                   unknown}:
##     :       :   :...installment_plan = bank: Yes (15.7/2.6)
##     :       :       installment_plan = stores: No (1.3/0.7)
##     :       :       installment_plan = none:
##     :       :       :...dependents > 1: No (15.8/5.3)
##     :       :           dependents <= 1:
##     :       :           :...installment_rate <= 1: No (13.3/5.2)
##     :       :               installment_rate > 1: Yes (67.9/19.4)
##     :       purpose = furniture:
##     :       :...installment_plan = stores: No (5.5)
##     :           installment_plan in {bank,none}:
##     :           :...other_debtors = guarantor: No (3.9)
##     :               other_debtors in {co-applicant,none}:
##     :               :...savings_balance in {> 1000 DM,unknown}: No (10.1/2.9)
##     :                   savings_balance in {101 - 500 DM,
##     :                   :                   501 - 1000 DM}: Yes (3.5)
##     :                   savings_balance = < 100 DM:
##     :                   :...amount <= 4473: No (66.2/30.1)
##     :                       amount > 4473: Yes (7)
##     checking_balance = unknown:
##     :...other_debtors = co-applicant: Yes (13.6/5.2)
##         other_debtors = guarantor: No (3.9)
##         other_debtors = none:
##         :...installment_plan = bank: Yes (50/21.1)
##             installment_plan in {none,stores}:
##             :...purpose in {car (used),domestic appliances,others,radio/tv,
##                 :           retraining}: No (101.9/8.4)
##                 purpose in {business,car (new),education,furniture,repairs}:
##                 :...amount > 7763: Yes (14.9/2)
##                     amount <= 7763:
##                     :...credit_history in {critical,
##                         :                  fully repaid this bank}: No (42.2/4.6)
##                         credit_history in {delayed,fully repaid,repaid}:
##                         :...savings_balance = > 1000 DM: Yes (8.2/2.6)
##                             savings_balance in {101 - 500 DM,501 - 1000 DM,
##                             :                   unknown}: No (28.8/8.4)
##                             savings_balance = < 100 DM:
##                             :...amount <= 1778: No (10.5)
##                                 amount > 1778: Yes (32.3/9.2)
```

```
##
## -----  Trial 3:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...foreign_worker = no: No (9.9)
## :   foreign_worker = yes:
## :   :...employment_length in {> 7 yrs,4 - 7 yrs}:
## :       :...dependents <= 1: No (112.3/11.8)
## :       :   dependents > 1:
## :       :   :...checking_balance = > 200 DM: Yes (5/0.5)
## :       :       checking_balance = unknown: No (34.4/11.6)
## :       employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}:
## :       :...other_debtors = co-applicant: Yes (12.8/3.5)
## :           other_debtors = guarantor: No (2.7)
## :           other_debtors = none:
## :           :...purpose in {business,repairs}: Yes (33.3/9.6)
## :               purpose in {car (used),domestic appliances,education,others,
## :               :           retraining}: No (36.8/9.2)
## :               purpose = car (new):
## :               :...housing in {for free,own}: No (28.7/9.2)
## :               :   housing = rent: Yes (5.6)
## :               purpose = furniture:
## :               :...job in {mangement self-employed,
## :               :   :       unskilled resident}: Yes (18.6/6.9)
## :               :   job in {skilled employee,
## :               :           unemployed non-resident}: No (18.3/2.3)
## :               purpose = radio/tv:
## :               :...job = unemployed non-resident: No (0)
## :                   job in {mangement self-employed,
## :                   :       unskilled resident}: Yes (16.8/5.6)
## :                   job = skilled employee:
## :                   :...amount <= 4057: No (28.2/1.5)
## :                       amount > 4057: Yes (9.1/3.5)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...credit_history in {fully repaid,fully repaid this bank}: Yes (64.1/18.1)
##     credit_history in {critical,delayed,repaid}:
##     :...other_debtors = co-applicant: Yes (23.3/9.6)
##         other_debtors = guarantor: No (27.7/9.1)
##         other_debtors = none:
##         :...purpose in {business,others}: No (31/7.6)
##             purpose in {domestic appliances,education,furniture,repairs,
##             :           retraining}: Yes (134/54.6)
##             purpose = car (used):
##             :...amount <= 8086: No (28.8/7.8)
##             :   amount > 8086: Yes (11.8/1.2)
##             purpose = car (new):
##             :...amount > 11054: Yes (6.5)
##             :   amount <= 11054:
##             :   :...personal_status = divorced male: Yes (7.5/2.4)
##             :       personal_status = married male: No (5.7/2.6)
##             :       personal_status = female:
##             :       :...amount <= 7418: Yes (26/5.7)
```

```
##             :       :   amount > 7418: No (4.4)
##             :       personal_status = single male:
##             :       :...months_loan_duration <= 42: No (55.8/16.6)
##             :           months_loan_duration > 42: Yes (2.9)
##             purpose = radio/tv:
##             :...foreign_worker = no: No (2.4)
##                 foreign_worker = yes:
##                 :...savings_balance in {> 1000 DM,501 - 1000 DM,
##                     :                   unknown}: No (22.5/7.8)
##                     savings_balance = 101 - 500 DM: Yes (10.8/1.1)
##                     savings_balance = < 100 DM:
##                     :...months_loan_duration > 39: Yes (6.1)
##                         months_loan_duration <= 39:
##                         :...amount > 3275: No (7.4)
##                             amount <= 3275:
##                             :...months_loan_duration <= 13: No (19.4/6.9)
##                                 months_loan_duration > 13: Yes (29/7.4)
##
## -----  Trial 4:  -----
##
## Decision tree:
##
## checking_balance in {> 200 DM,unknown}:
## :...purpose in {car (used),domestic appliances,education,others,radio/tv,
## :   :           retraining}: No (169.6/44.2)
## :   purpose = repairs: Yes (8.9/4.1)
## :   purpose = business:
## :   :...employment_length in {> 7 yrs,1 - 4 yrs,4 - 7 yrs}: No (29.9/6.4)
## :   :   employment_length in {0 - 1 yrs,unemployed}: Yes (12.5/1.1)
## :   purpose = car (new):
## :   :...installment_plan in {bank,stores}: Yes (20.3/7)
## :   :   installment_plan = none:
## :   :   :...amount <= 11760: No (51.5/14.6)
## :   :       amount > 11760: Yes (2.8)
## :   purpose = furniture:
## :   :...credit_history in {critical,fully repaid,
## :       :                  fully repaid this bank}: No (15.2)
## :       credit_history in {delayed,repaid}:
## :       :...other_debtors = guarantor: No (0)
## :           other_debtors = co-applicant: Yes (3.9)
## :           other_debtors = none:
## :           :...months_loan_duration <= 30: No (36.8/11.6)
## :               months_loan_duration > 30: Yes (4.4/0.5)
## checking_balance in {< 0 DM,1 - 200 DM}:
## :...savings_balance in {> 1000 DM,unknown}:
##     :...credit_history in {critical,delayed,fully repaid}: No (23.5/3.2)
##     :   credit_history = fully repaid this bank: Yes (5.2/2.1)
##     :   credit_history = repaid:
##     :   :...amount <= 5771: No (41.4/8.8)
##     :       amount > 5771: Yes (15/2.2)
##     savings_balance in {< 100 DM,101 - 500 DM,501 - 1000 DM}:
##     :...months_loan_duration > 42: Yes (37.3/8.8)
##         months_loan_duration <= 42:
##         :...purpose in {car (used),domestic appliances,others,
```

```
##             :           retraining}: No (47.4/17.5)
##             purpose in {education,repairs}: Yes (34.4/15)
##             purpose = business:
##             :...months_loan_duration <= 18: No (10.2)
##             :   months_loan_duration > 18: Yes (20.1/5.7)
##             purpose = car (new):
##             :...other_debtors in {co-applicant,guarantor}: Yes (15/3.5)
##             :   other_debtors = none:
##             :   :...installment_rate <= 3:
##             :       :...residence_history <= 1: Yes (5.9)
##             :       :   residence_history > 1: No (44.1/15.8)
##             :       installment_rate > 3:
##             :       :...amount <= 609: No (4.6)
##             :           amount > 609: Yes (37.7/7.8)
##             purpose = radio/tv:
##             :...foreign_worker = no: No (2.9)
##             :   foreign_worker = yes:
##             :   :...employment_length in {> 7 yrs,4 - 7 yrs}: No (26.5/8.4)
##             :       employment_length in {0 - 1 yrs,unemployed}: Yes (33.9/12.6)
##             :       employment_length = 1 - 4 yrs:
##             :       :...months_loan_duration <= 11: No (6.2)
##             :           months_loan_duration > 11: Yes (28.4/10.8)
##             purpose = furniture:
##             :...installment_plan = stores: No (4.9)
##                 installment_plan in {bank,none}:
##                 :...credit_history in {critical,fully repaid}: No (32.9/13.4)
##                     credit_history in {delayed,
##                     :                  fully repaid this bank}: Yes (8.1/1.1)
##                     credit_history = repaid:
##                     :...checking_balance = 1 - 200 DM: Yes (17.6/6)
##                         checking_balance = < 0 DM:
##                         :...months_loan_duration <= 15: No (16.5/2)
##                             months_loan_duration > 15: Yes (24.3/11)
##
## -----  Trial 5:  -----
##
## Decision tree:
##
## foreign_worker = no: No (23.7/4.5)
## foreign_worker = yes:
## :...checking_balance = < 0 DM:
##     :...job = mangement self-employed:
##     :   :...installment_rate <= 1: Yes (4.3)
##     :   :   installment_rate > 1: No (37.5/9.2)
##     :   job in {skilled employee,unemployed non-resident,unskilled resident}:
##     :   :...months_loan_duration <= 8: No (11.8/1.4)
##     :       months_loan_duration > 8:
##     :       :...purpose in {business,furniture,others}: No (73.2/31.3)
##     :           purpose in {car (new),car (used),domestic appliances,education,
##     :           :           repairs,retraining}: Yes (102.3/33)
##     :           purpose = radio/tv:
##     :           :...employment_length = > 7 yrs: No (4)
##     :               employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,
##     :                                     unemployed}: Yes (33.9/9.2)
```

```
##     checking_balance = > 200 DM:
##     :...dependents > 1: Yes (7.3/0.9)
##     :   dependents <= 1:
##     :   :...age > 39: No (14.2)
##     :       age <= 39:
##     :       :...age <= 24: No (7.4)
##     :           age > 24:
##     :           :...installment_plan = bank: No (3.7)
##     :               installment_plan in {none,stores}: Yes (31.1/8.4)
##     checking_balance = 1 - 200 DM:
##     :...employment_length = 4 - 7 yrs: No (42/8.4)
##     :   employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,unemployed}:
##     :   :...amount > 12204: Yes (12.4)
##     :       amount <= 12204:
##     :       :...dependents > 1: Yes (24.6/7.7)
##     :           dependents <= 1:
##     :           :...housing = for free: No (20.7/4.2)
##     :               housing = rent:
##     :               :...savings_balance = 101 - 500 DM: Yes (11.1)
##     :               :   savings_balance in {< 100 DM,> 1000 DM,501 - 1000 DM,
##     :               :   :                   unknown}:
##     :               :   :...employment_length in {> 7 yrs,
##     :               :       :                     0 - 1 yrs}: Yes (23.4/7.5)
##     :               :       employment_length in {1 - 4 yrs,
##     :               :                             unemployed}: No (8.3)
##     :               housing = own:
##     :               :...residence_history <= 1: No (36.1/6.9)
##     :                   residence_history > 1:
##     :                   :...savings_balance = unknown: No (12.9/1)
##     :                       savings_balance in {< 100 DM,> 1000 DM,
##     :                       :                   101 - 500 DM,501 - 1000 DM}:
##     :                       :...job in {mangement self-employed,
##     :                           :       unemployed non-resident}: Yes (13.4/1.7)
##     :                           job in {skilled employee,unskilled resident}: [S1]
##     checking_balance = unknown:
##     :...credit_history in {critical,fully repaid this bank}: No (97.7/24.4)
##         credit_history = fully repaid: Yes (7/3.5)
##         credit_history = delayed:
##         :...installment_rate <= 3: No (14.3/2.7)
##         :   installment_rate > 3: Yes (23.2/5.4)
##         credit_history = repaid:
##         :...savings_balance = 101 - 500 DM: No (9.8)
##             savings_balance in {< 100 DM,> 1000 DM,501 - 1000 DM,unknown}:
##             :...existing_credits > 1: Yes (24.7/8.8)
##                 existing_credits <= 1:
##                 :...age > 41: No (14.8/1.6)
##                     age <= 41:
##                     :...residence_history <= 1: No (4.3)
##                         residence_history > 1:
##                         :...savings_balance in {> 1000 DM,
##                             :                   unknown}: No (16.1/1.8)
##                             savings_balance in {< 100 DM,501 - 1000 DM}:
##                             :...personal_status in {divorced male,
##                                 :                   married male}: No (3.8)
```

```
##                                 personal_status in {female,single male}:
##                                 :...telephone = yes: No (13.8/5.3)
##                                     telephone = none: [S2]
##
## SubTree [S1]
##
## employment_length in {> 7 yrs,unemployed}: No (14.3/2.9)
## employment_length = 0 - 1 yrs: Yes (17.4/6.5)
## employment_length = 1 - 4 yrs:
## :...months_loan_duration > 22: Yes (7.2/0.4)
##     months_loan_duration <= 22:
##     :...age <= 55: No (26.1/5.7)
##         age > 55: Yes (5.6/0.8)
##
## SubTree [S2]
##
## job = unemployed non-resident: Yes (0)
## job = mangement self-employed: No (2.4)
## job in {skilled employee,unskilled resident}:
## :...months_loan_duration <= 30: Yes (35.3/7.8)
##     months_loan_duration > 30: No (2.9)
##
## -----  Trial 6:  -----
##
## Decision tree:
##
## amount > 6419: Yes (134.8/52.8)
## amount <= 6419:
## :...months_loan_duration <= 7: No (54.4/10.7)
##     months_loan_duration > 7:
##     :...checking_balance = unknown:
##         :...installment_plan = stores: No (14.1/6.9)
##         :   installment_plan = bank:
##         :   :...age > 43: No (8.3)
##         :   :   age <= 43:
##         :   :   :...age <= 31: No (11.9/2.9)
##         :   :       age > 31: Yes (20.4/3.4)
##         :   installment_plan = none:
##         :   :...credit_history in {critical,fully repaid,
##         :       :                   fully repaid this bank}: No (43.1)
##         :       credit_history in {delayed,repaid}:
##         :       :...residence_history <= 1: Yes (10.4/2.6)
##         :           residence_history > 1:
##         :           :...savings_balance in {101 - 500 DM,
##         :               :                   unknown}: No (19.9)
##         :               savings_balance in {< 100 DM,> 1000 DM,501 - 1000 DM}:
##         :               :...other_debtors = co-applicant: Yes (4.1/0.7)
##         :                   other_debtors = guarantor: No (0.3)
##         :                   other_debtors = none:
##         :                   :...age > 29: No (35.2/6.9)
##         :                       age <= 29:
##         :                       :...installment_rate <= 3: No (17.4/5.7)
##         :                           installment_rate > 3: Yes (18.5/3.6)
##         checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
```

```
##         :...residence_history <= 1:
##             :...installment_plan in {bank,stores}: No (16.5)
##             :   installment_plan = none:
##             :   :...other_debtors in {co-applicant,guarantor}: No (3.5)
##             :       other_debtors = none:
##             :       :...job = mangement self-employed: No (5)
##             :           job in {unemployed non-resident,
##             :           :       unskilled resident}: Yes (9.7/0.9)
##             :           job = skilled employee:
##             :           :...housing in {for free,rent}: Yes (7.8/1.2)
##             :               housing = own:
##             :               :...checking_balance = < 0 DM: Yes (8.8/1.9)
##             :                   checking_balance in {> 200 DM,
##             :                                        1 - 200 DM}: No (32.2/5.5)
##             residence_history > 1:
##             :...installment_rate > 2:
##                 :...job = unemployed non-resident: No (5.4)
##                 :   job in {mangement self-employed,skilled employee,
##                 :   :       unskilled resident}:
##                 :   :...telephone = none:
##                 :       :...installment_plan in {bank,stores}: Yes (40.7/7.7)
##                 :       :   installment_plan = none:
##                 :       :   :...personal_status in {divorced male,
##                 :       :       :                   married male}: No (20.3/8.8)
##                 :       :       personal_status = female: Yes (44.5/13.9)
##                 :       :       personal_status = single male: [S1]
##                 :       telephone = yes:
##                 :       :...other_debtors in {co-applicant,
##                 :           :                 guarantor}: No (12.3/0.9)
##                 :           other_debtors = none:
##                 :           :...savings_balance in {> 1000 DM,101 - 500 DM,
##                 :               :                   unknown}: No (24.5/5.1)
##                 :               savings_balance = 501 - 1000 DM: Yes (6.3/2.7)
##                 :               savings_balance = < 100 DM: [S2]
##                 installment_rate <= 2:
##                 :...credit_history = delayed: No (9)
##                     credit_history in {critical,fully repaid,
##                     :                  fully repaid this bank,repaid}:
##                     :...housing in {for free,rent}: No (49.7/15.4)
##                         housing = own:
##                         :...foreign_worker = no: Yes (2.6/0.3)
##                             foreign_worker = yes:
##                             :...installment_rate <= 1: No (15.2/4.8)
##                                 installment_rate > 1:
##                                 :...months_loan_duration > 36: Yes (5)
##                                     months_loan_duration <= 36: [S3]
##
## SubTree [S1]
##
## savings_balance in {> 1000 DM,501 - 1000 DM}: Yes (5.9/0.7)
## savings_balance in {101 - 500 DM,unknown}: No (18.4/3.8)
## savings_balance = < 100 DM:
## :...property in {building society savings,real estate}: No (37.6/13)
##     property in {other,unknown/none}: Yes (22.4/4.7)
```

```
##
## SubTree [S2]
##
## credit_history in {delayed,fully repaid,fully repaid this bank}: Yes (11.2)
## credit_history in {critical,repaid}:
## :...job = mangement self-employed: No (10.6/2)
##     job in {skilled employee,unskilled resident}: Yes (28.2/9.4)
##
## SubTree [S3]
##
## other_debtors = co-applicant: Yes (4.6/1.9)
## other_debtors = guarantor: No (2.7)
## other_debtors = none:
## :...amount <= 3416: Yes (33.8/14.4)
##     amount > 3416: No (11.6/1.3)
##
## -----  Trial 7:  -----
##
## Decision tree:
##
## credit_history in {fully repaid,fully repaid this bank}:
## :...property in {building society savings,unknown/none}: Yes (47.1/12.3)
## :   property in {other,real estate}:
## :   :...savings_balance in {< 100 DM,unknown}: Yes (41.9/17.7)
## :       savings_balance in {> 1000 DM,101 - 500 DM,
## :                           501 - 1000 DM}: No (15.2/1.3)
## credit_history in {critical,delayed,repaid}:
## :...checking_balance in {> 200 DM,unknown}: No (304.8/95)
##     checking_balance in {< 0 DM,1 - 200 DM}:
##     :...property = real estate:
##         :...savings_balance in {> 1000 DM,101 - 500 DM,501 - 1000 DM,
##         :   :                   unknown}: No (28.1/3.9)
##         :   savings_balance = < 100 DM:
##         :   :...age > 33: No (43.4/7.3)
##         :       age <= 33:
##         :       :...amount <= 1217: No (17.7/3.7)
##         :           amount > 1217: Yes (33/7.9)
##         property in {building society savings,other,unknown/none}:
##         :...amount <= 959: Yes (37.7/8.2)
##             amount > 959:
##             :...dependents > 1: No (50/16.5)
##                 dependents <= 1:
##                 :...months_loan_duration > 27:
##                     :...job = unemployed non-resident: No (2.7)
##                     :   job = unskilled resident: Yes (6.5)
##                     :   job in {mangement self-employed,skilled employee}:
##                     :   :...credit_history = delayed: No (17.1/7.4)
##                     :       credit_history in {critical,repaid}:
##                     :       :...residence_history <= 1: No (6.3/1.1)
##                     :           residence_history > 1: Yes (49.2/14.5)
##                     months_loan_duration <= 27:
##                     :...personal_status = married male: Yes (16.7/4.4)
##                         personal_status in {divorced male,female,single male}:
##                         :...credit_history in {critical,
```

```
##                               :                  delayed}: No (64.6/13.5)
##                               credit_history = repaid:
##                               :...amount > 10222: Yes (5.8)
##                                   amount <= 10222:
##                                   :...age > 54: No (10.3/1.3)
##                                       age <= 54:
##                                       :...age <= 31: No (66.7/19.4)
##                                           age > 31: Yes (32/8.5)
##
## -----  Trial 8:  -----
##
## Decision tree:
##
## housing in {for free,rent}:
## :...purpose in {business,domestic appliances,repairs,
## :   :           retraining}: Yes (31.2/6.8)
## :   purpose in {education,others}: No (30.1/11.4)
## :   purpose = car (used):
## :   :...amount <= 11054: No (40.8/12.4)
## :   :   amount > 11054: Yes (5.8)
## :   purpose = car (new):
## :   :...employment_length = unemployed: No (8.8/0.7)
## :   :   employment_length in {> 7 yrs,0 - 1 yrs,1 - 4 yrs,4 - 7 yrs}:
## :   :   :...months_loan_duration <= 9: No (3.9)
## :   :       months_loan_duration > 9: Yes (59.7/16.1)
## :   purpose = furniture:
## :   :...credit_history = delayed: No (0)
## :   :   credit_history in {fully repaid,fully repaid this bank}: Yes (6.1)
## :   :   credit_history in {critical,repaid}:
## :   :   :...job = mangement self-employed: Yes (12.6/3.2)
## :   :       job in {skilled employee,unemployed non-resident,
## :   :               unskilled resident}: No (45.9/15.3)
## :   purpose = radio/tv:
## :   :...job in {mangement self-employed,unemployed non-resident}: No (10.8)
## :       job in {skilled employee,unskilled resident}:
## :       :...employment_length = > 7 yrs: No (3.9)
## :           employment_length in {0 - 1 yrs,1 - 4 yrs,4 - 7 yrs,
## :                                 unemployed}: Yes (28.6/8)
## housing = own:
## :...purpose in {car (used),domestic appliances,repairs,
##     :           retraining}: No (72.6/14.3)
##     purpose in {education,others}: Yes (27.1/11.7)
##     purpose = business:
##     :...savings_balance = > 1000 DM: Yes (6.4/3.1)
##     :   savings_balance in {101 - 500 DM,501 - 1000 DM,unknown}: No (23.1/4)
##     :   savings_balance = < 100 DM:
##     :   :...amount > 7596: Yes (6.3)
##     :       amount <= 7596:
##     :       :...installment_plan = bank: No (5.7)
##     :           installment_plan in {none,stores}:
##     :           :...telephone = none: No (16.9/4.3)
##     :               telephone = yes: Yes (23.1/8.5)
##     purpose = car (new):
##     :...foreign_worker = no: No (5.2)
```

```
##     :   foreign_worker = yes:
##     :   :...installment_rate <= 2:
##     :       :...existing_credits > 3: Yes (2.2)
##     :       :   existing_credits <= 3:
##     :       :   :...age <= 23: Yes (3.9)
##     :       :       age > 23: No (42.6/10.5)
##     :       installment_rate > 2:
##     :       :...installment_plan in {bank,stores}: Yes (13.1)
##     :           installment_plan = none:
##     :           :...checking_balance in {< 0 DM,> 200 DM,
##     :               :                    1 - 200 DM}: Yes (57.9/24.1)
##     :               checking_balance = unknown: No (8.7)
##     purpose = furniture:
##     :...installment_plan = stores: No (9)
##     :   installment_plan in {bank,none}:
##     :   :...credit_history = fully repaid: No (5.3)
##     :       credit_history = fully repaid this bank: Yes (4.1)
##     :       credit_history in {critical,delayed,repaid}:
##     :       :...telephone = none:
##     :           :...months_loan_duration <= 15: No (26.7/7.9)
##     :           :   months_loan_duration > 15: Yes (42.8/12.8)
##     :           telephone = yes:
##     :           :...job = mangement self-employed: Yes (11.9/4.7)
##     :               job in {skilled employee,unemployed non-resident,
##     :                       unskilled resident}: No (19.9/1.2)
##     purpose = radio/tv:
##     :...checking_balance = unknown: No (49.7/5.9)
##         checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
##         :...months_loan_duration > 36: Yes (13.4/1.5)
##             months_loan_duration <= 36:
##             :...other_debtors = co-applicant: Yes (2.3)
##                 other_debtors = guarantor: No (12.1)
##                 other_debtors = none:
##                 :...employment_length in {> 7 yrs,1 - 4 yrs,
##                     :                     4 - 7 yrs}: No (61.4/17.7)
##                     employment_length in {0 - 1 yrs,unemployed}: Yes (29.2/8.7)
##
## -----  Trial 9:  -----
##
## Decision tree:
##
## checking_balance = unknown:
## :...employment_length in {> 7 yrs,4 - 7 yrs}: No (89.1/5.2)
## :   employment_length in {0 - 1 yrs,1 - 4 yrs,unemployed}:
## :   :...installment_plan in {bank,stores}:
## :       :...other_debtors in {co-applicant,guarantor}: No (3.9)
## :       :   other_debtors = none:
## :       :   :...residence_history <= 1: No (3.2)
## :       :       residence_history > 1:
## :       :       :...purpose in {business,car (new),car (used),
## :       :           :           domestic appliances,furniture,others,repairs,
## :       :           :           retraining}: Yes (30.3/4.6)
## :       :           purpose in {education,radio/tv}: No (4.7)
## :       installment_plan = none:
```

```
## :       :...other_debtors = co-applicant: Yes (7.9/1.4)
## :           other_debtors in {guarantor,none}:
## :           :...months_loan_duration <= 16: No (31)
## :               months_loan_duration > 16:
## :               :...property in {building society savings,
## :                   :            unknown/none}: No (15.6)
## :                   property in {other,real estate}:
## :                   :...credit_history in {critical,fully repaid,
## :                       :                  fully repaid this bank}: No (9.7)
## :                       credit_history in {delayed,repaid}: Yes (27.6/10.8)
## checking_balance in {< 0 DM,> 200 DM,1 - 200 DM}:
## :...savings_balance in {> 1000 DM,501 - 1000 DM,unknown}:
##     :...savings_balance = > 1000 DM: No (26.7/6.9)
##     :   savings_balance in {501 - 1000 DM,unknown}:
##     :   :...installment_rate > 3:
##     :       :...residence_history <= 3: Yes (30.5/9.5)
##     :       :   residence_history > 3: No (24.2/2.5)
##     :       installment_rate <= 3:
##     :       :...housing = for free: Yes (5.9/1.6)
##     :           housing = rent: No (8.7)
##     :           housing = own:
##     :           :...age <= 23: Yes (4.8)
##     :               age > 23: No (30.2/1.8)
##     savings_balance in {< 100 DM,101 - 500 DM}:
##     :...months_loan_duration > 47: Yes (31.5/6.1)
##         months_loan_duration <= 47:
##         :...other_debtors = co-applicant: Yes (27.6/13.5)
##             other_debtors = guarantor:
##             :...installment_plan = bank: Yes (11.7/4.3)
##             :   installment_plan in {none,stores}: No (20.5/3)
##             other_debtors = none:
##             :...credit_history in {fully repaid,
##                 :                  fully repaid this bank}: Yes (51.4/17.4)
##                 credit_history = delayed:
##                 :...installment_rate <= 1: No (6.1)
##                 :   installment_rate > 1:
##                 :   :...savings_balance = < 100 DM: Yes (23/8)
##                 :       savings_balance = 101 - 500 DM: No (10.3/2.9)
##                 credit_history = critical:
##                 :...savings_balance = 101 - 500 DM: No (7.4/1)
##                 :   savings_balance = < 100 DM:
##                 :   :...personal_status = divorced male: Yes (9/1.3)
##                 :       personal_status in {female,married male,single male}:
##                 :       :...telephone = yes: No (31.3/6.8)
##                 :           telephone = none: [S1]
##                 credit_history = repaid:
##                 :...installment_rate <= 1: Yes (31/9.9)
##                     installment_rate > 1:
##                     :...job = unemployed non-resident: Yes (1.5)
##                         job = mangement self-employed:
##                         :...amount <= 7582: No (26.9/3.8)
##                         :   amount > 7582: Yes (4.2)
##                         job in {skilled employee,unskilled resident}:
##                         :...foreign_worker = no: No (2.2)
```

```
##                               foreign_worker = yes:
##                               :...installment_plan = stores: No (4.4)
##                                   installment_plan in {bank,none}:
##                                   :...installment_rate > 3: Yes (85.2/32.7)
##                                       installment_rate <= 3: [S2]
##
## SubTree [S1]
##
## property in {building society savings,real estate,unknown/none}: No (37.7/10)
## property = other: Yes (7.9/0.2)
##
## SubTree [S2]
##
## personal_status in {married male,single male}: No (50.5/13.1)
## personal_status in {divorced male,female}:
## :...existing_credits > 1: Yes (3.9)
##     existing_credits <= 1:
##     :...savings_balance = < 100 DM: Yes (36.5/13.2)
##         savings_balance = 101 - 500 DM: No (3.2)
##
##
## Evaluation on training data (900 cases):
##
## Trial        Decision Tree
## -----      ----------------
##    Size      Errors
##
##    0     42  136(15.1%)
##    1     27  189(21.0%)
##    2     33  215(23.9%)
##    3     36  210(23.3%)
##    4     36  180(20.0%)
##    5     42  199(22.1%)
##    6     44  216(24.0%)
##    7     21  190(21.1%)
##    8     41  211(23.4%)
##    9     41  181(20.1%)
## boost             45( 5.0%)   <<
##
##
##     (a)   (b)    <-classified as
##    ----  ----
##     628     7    (a): class No
##      38   227    (b): class Yes
##
##
##  Attribute usage:
##
##  100.00% checking_balance
##  100.00% credit_history
##  100.00% purpose
##  100.00% amount
##  100.00% housing
##  100.00% foreign_worker
```

```
##   99.56% other_debtors
##   96.78% months_loan_duration
##   87.56% employment_length
##   85.89% installment_plan
##   83.89% savings_balance
##   74.22% residence_history
##   70.11% dependents
##   70.00% property
##   68.56% installment_rate
##   68.00% job
##   51.33% personal_status
##   49.44% age
##   47.00% telephone
##   23.22% existing_credits
##
##
## Time: 0.1 secs
```

Checking enhanced accuracy

```
credit_boost_pred10 <- predict(credit_boost10, credit_test)
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 |        59 |         6 |        65 |
##                |     0.590 |     0.060 |           |
## ---------------|-----------|-----------|-----------|
##            Yes |        17 |        18 |        35 |
##                |     0.170 |     0.180 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        76 |        24 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

Accuracy after boosting is 77% which is 9% more than the previous approach

Assigning weights to the miss classifications

```
# making cost matrix
matrix_dimensions <- list(c("no", "yes"), c("no", "yes"))
names(matrix_dimensions) <- c("predicted", "actual")
error_cost <- matrix(c(0, 1, 4, 0), nrow = 2)
error_cost
```

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

Performance evaluation with cost matrix

```
credit_cost <- C5.0(credit_train[-17], label, 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, dr
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  100
##
##
##                | predicted default
## actual default |        No |       Yes | Row Total |
## ---------------|-----------|-----------|-----------|
##             No |        43 |        22 |        65 |
##                |     0.430 |     0.220 |           |
## ---------------|-----------|-----------|-----------|
##            Yes |         8 |        27 |        35 |
##                |     0.080 |     0.270 |           |
## ---------------|-----------|-----------|-----------|
##   Column Total |        51 |        49 |       100 |
## ---------------|-----------|-----------|-----------|
##
##
```

Here accuracy comes out to be 70% which is less than boostiong method also this error cost method has high error rate than boosting method. However, if assigned weights are accurate, higher accuracy could be acheived.

Problem 2.

Step 2 - Exploring and preparing the data

```
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 ...
```

```
# As veil_type column is not significant thus  neglected
mushrooms$veil_type <- NULL
table(mushrooms$type)
```

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

Step 3 - training the model on the data

```
library(RWeka)
```

```
## Warning: package 'RWeka' was built under R version 3.6.2
```

```
# using RWeka package to create decision tree with type column in dataset
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 the model Performance

```
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 JRip function of RWeka package for decision tree
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
```

Problem 3.

KNN Algorithm

KNN is a simple algorithm that treats the features as coordinates in a multidimensional feature space. It also assumes similar cases are within close vicinity, while different cases are more distant to each other. Similarity between Instances are compared using Euclidean distance, thus there is a need of normalization and standardization. It is a non- parametric model I.e. no parameters are learned about the data. No learning of the model is required and all of the work happens at the time a prediction is requested. KNN algorithm usually performs better with small datasets with numeric features. This simple algorithm is useful when there's no need to build a model, or tune parameters and make additional assumptions. As training phase occur very rapidly, the downside is that the process of making predictions tends to be relatively slow in comparison to training. Another disadvantage of kNN is its computational cost for new classifications, particularly with larger values of k. KNN tends to overfits the data. It is not an ideal algorithm to implement on high dimensional data, KNN is also sensitive to noisy data, missing values and outliers.

Naive Bayes

This classification algorithm is based on Bayes Theorem of conditional probability. Naive Bayes assumes that all of the features in the dataset are equally important and independent i.e. presence of a particular feature in a class is unrelated to the presence of any other feature. This is a simple and fast algorithm that does well with the case of categorical input variables compared to numerical variables. It can handle data with noisy and missing values. There is no need for normalization and standardization of data as it relies on frequency of data. It also perform well with high dimensional data and multi class prediction. The major disadvantage of this algorithm is that it relies on assumption of equally important and independent features. It's not ideal to use with numeric features. Another problem with this algorithm is "Zero-Frequency" I.e if a categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. To overcome this a smoothing technique is used.

C5.0 Decision Trees

This algorithm works on divide and conquer strategy, decision trees are generated by splitting the data, attributes of the data are chosen that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the information gain (difference in entropy), Gini index, etc. A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders. Decisi rees work well with with large datasets and can handle noisy, outlier-prone data with relatively high accuracy. Decision trees are well-suited for data with a few highly relevant variables. This algorithm doesn't require scaling and normalization of data, also missing values in the data doesn't affect the process of building decision tree to any considerable extent. The limitation of this algorithm is that a small change in the data can cause a large change in the structure of the decision tree causing instability. For a decision tree sometimes calculation can go far more complex compared to other algorithms thus, it require higher time to train the model. Decision tree training is relatively expensive as complexity and time taken is more. The model can be easily overfitted or underfitted although this could be adjusted using pruning.

1R Algorithm

Rule-based learners use a "separate and conquer" technique to identify rules that cover all instances in the dataset. 1R is a simple algorithm that utilizes one feature to inform rules and build decision tree. The error rate for the rule based on each feature is calculated and the rule with the fewest errors is chosen as the one rule. This algorithm performs well despite its simple approach. It works well on predominantly categorical data. 1R generates a single, easy-to-understand, human-readable rules to generate decision trees. Its disadvantage is that it uses only a single feature when considering rules for a complex data with large feature values.

Ripper Algorithm

This algorithm is based on separate and conquer approach. This algorithm creates much more complex rules than 1R algorithm, as more than one features are considered and rules with multiple antecedents are also created. Incremental pruning (tuning the feature vales) for error reduction is done for this algorithm. This improves the algorithm's ability to model complex data. This algorithm is not ideal to work with numeric datasets although it can handle large noisy data.

Problem 4.

A prediction model that is composed of a set of models is called a model ensemble. Each model makes prediction independently of other models in the ensemble. Given a large population of independent models, an ensemble can be very accurate even if the individual models in the ensemble perform only marginally better than random guessing. Different models from same dataset is built by inducing each model using a modified version of dataset. Final predictions are made by aggregating the predictions of different models in the ensemble. One of the benefits of using ensembles is that they may allow user to spend less time in pursuit of a single best model. The two standard approaches for creating ensemble models are:

Boosting : In this technique more weights are assigned to the misclassifications in each iteration, so that each iterative model pays more attention to instances misclassified by the previous model. Models are iteratively created and added to the ensemble and stops when a predefined number of models have been added. Once models are created, the ensemble makes predictions using a weighted aggregate of the predictions made by individual models.

Bagging : In this approach each model in the ensemble is trained on a random sample of the dataset and sampling is done with replacement so that samples are different and models trained are also different. This technique works well with decision trees as overfitting of models could be avoided. In bagging each model is constructed independently and each model is assigned equal weight is giving out the final prediction.