Subject: 810198358 امیرمهدی انصاری پور

Date:

Page:

4

[illegible]) یکی از drawback های احتمالی busheir tree (که غالبا از الگوریتم C45 بدست می آید)، این است که به تعداد برگ های بیشتری نسبت به binary tree منجر میشود. درنتیجه تعداد قواعد تقسیم زیادتر شده که معادل overfitting می باشد.

برای ساختن دست تقسیم، در ابتدا باید Age attribute را Binning کنیم. برای این کار داریم،

$25 \leqslant Age \leqslant 26 \Rightarrow low$

$30 \leqslant Age \leqslant 35 \Rightarrow mid$

$40 \leqslant Age \Rightarrow high$

$$\Phi(split|t) = 2P_L P_R \sum_{j=1}^{\#class} |P(j|t_L) - P(j|t_R)|$$

5) الخوارزمية CART : بدك شجر هدف Salary :

شجر age :

Age = {low, mid} {high} Split 1

Age = {low} {mid, high} Split 2

Age = {low, high} {mid} Split 3

Split 1 : {high} = $\overset{L3}{R1}, \overset{L4}{R_7}, \overset{L3}{R_8}, \overset{L2}{R_{10}}$

{low, mid} = $\underset{L1}{R_2}, \underset{L2}{R_3}, \underset{L3}{R_4}, \underset{L4}{R_5}, \underset{L3}{R_6}, \underset{L2}{R_9}, \underset{L1}{R_{11}}$

$P_L = 4/_{11}$ $P_R = 7/_{11}$

$$\sum_{j=1}^{\#class} |P(j|t_L) - P(j|t_R)| = \overset{8/28}{|0 - 2/_7|} + \overset{\frac{1}{28}}{|1/_4 - 2/_7|} +$$

$$\underset{\frac{6}{28}}{|2/_4 - 2/_7|} + \underset{3/28}{|1/_4 - 1/_7|}$$

$14$
$-8$

$$\Rightarrow \boxed{\Phi(split1|t) = 0.46 * 0.64 = 0.29}$$

Split 2 = {low} = $\overset{L1}{R_2}, \overset{L4}{R_4}, \overset{L3}{R_6}, \overset{L1}{R_{11}}$

{high, mid} = $\underset{L3}{R_1}, \underset{L2}{R_3}, \underset{L4}{R_5}, \underset{L4}{R_7}, \underset{L3}{R_8}, \underset{L2}{R_9}, \underset{L2}{R_{10}}$

$P_L = 4/_{11}$ $P_R = 7/_{11}$

$$\sum^{\#} |P(j|t_L) - P(j|t_R)| = \overset{\frac{14}{28}}{|2/_4 - 0|} + \overset{\frac{12}{28}}{|0 - 3/_7|} + \overset{\frac{1}{28}}{|1/_4 - 2/_7|}$$

$$+ \underset{1/28}{|1/_4 - 2/_7|} = 1$$

$$\Rightarrow \boxed{\Phi(\text{split}\,2|t) = 0.46}$$

$$\text{Split 3}: \quad P_L = \overset{L2}{R_3}, \overset{L4}{R_5}, \overset{L2}{R_9} = 3/11$$

$$P_R = \overset{L3}{R_1}, \overset{L1}{R_2}, \overset{L3}{R_4}, \overset{L3}{R_6}, \overset{L4}{R_7}, \overset{L3}{R_8}, \overset{L2}{R_{10}}, \overset{L1}{R_{11}} = 8/11$$

$$\sum_{i=1}^{\#\text{class}} |P(i|t_L) - P(i|t_R)| = \overset{6/24}{\left|0 - 2/8\right|} + \overset{13/24}{\left|2/3 - 1/8\right|}$$

$$+ \underset{4/24}{\left|0/3 - 4/8\right|} + \underset{5/24}{\left|1/3 - 1/8\right|} =$$

$$\Rightarrow \boxed{\Phi(\text{split}\,3|t) = 0.46}$$

Split 1 : {male} , {female} = Gender [illegible]

$$\text{male} = \overset{L1}{R_2}, \overset{L2}{R_3}, \overset{L3}{R_4}, \overset{L3}{R_6}, \overset{L2}{R_9}, \overset{L1}{R_{11}} \Rightarrow P_L = 6/11$$

$$\text{female} = \underset{L3}{R_1}, \underset{L4}{R_5}, \underset{L4}{R_7}, \underset{L3}{R_8}, \underset{L2}{R_{10}} \Rightarrow P_R = 5/11$$

$$\sum_{i=1}^{\#\text{class}} |P(i|t_L) - P(i|t_R)| = \overset{10/30}{\left|2/6 - 0\right|} + \overset{4/30}{\left|2/6 - 1/5\right|}$$

$$+ \underset{2/30}{\left|2/6 - 2/5\right|} + \underset{12/30}{\left|0 - 2/5\right|}$$

$$= 28/30 \Rightarrow \boxed{\Phi(\text{split}\,1|t) = 0.46}$$

: occupation شغل

split1 = {Ser}, {Man, Sal, Staf}

Split2 = {Manu}, {Ser, Sal, Staf}

Split3 = {Sal}, {Ser, Man, Staf}

Split4 = {Staf}, {Serv, Manu, Sal}

Split I : $P_L = 3/11$ $P_R = 8/11$

$$Q(\text{split I}|t) = \left|\overset{5/24}{\cancel{\tfrac{1}{3} - \tfrac{1}{8}}}\right| + \left|\overset{2/24}{\cancel{\tfrac{1}{3} - \tfrac{2}{8}}}\right| + \left|\underset{1/24}{\cancel{\tfrac{1}{3} - \tfrac{3}{8}}}\right| + \left|\underset{6/24}{\cancel{0 - \tfrac{2}{8}}}\right|$$

$\boxed{\Phi(\text{split1}|t) = 0.23}$

Split 2 : $P_L = 4/11$ $P_R = 7/11$

$$Q(\text{split2}|t) = \left|\overset{8}{\cancel{2/7}} - 0\right| + \left|\overset{12}{\cancel{3/7}} - 0\right| + \left|\overset{6}{\cancel{2/4 - 2/7}}\right| + \left|\overset{14}{\cancel{2/4}}\right|$$

$\Rightarrow \boxed{\Phi(\text{split 2}|t) = 0.66}$

Split 3 : $P_L = 2/11$ $P_R = 9/11$

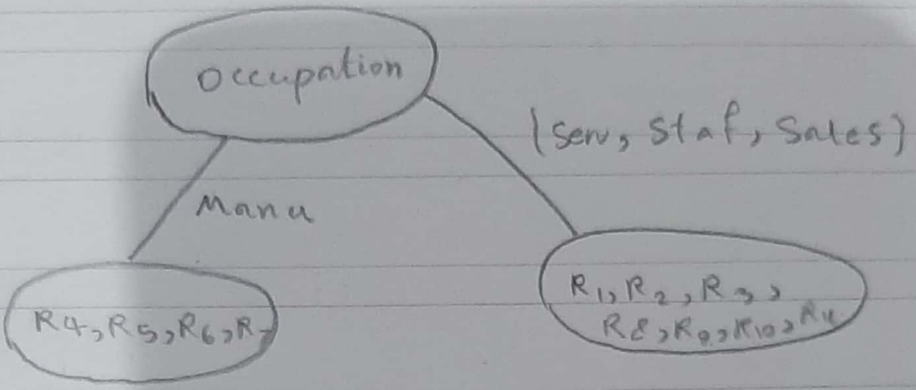
$$Q(\text{split3}|t) = \left|\overset{4}{\cancel{2/9}}\right| + \left|\overset{5}{\cancel{1/2 - 2/9}}\right| + \left|\overset{3}{\cancel{1/2 - 3/9}}\right| + \left|\overset{4}{\cancel{2/9}}\right|$$

$\boxed{\Phi(\text{split 3}|t) = 0.26}$

Split 4 : $P_L = 2/11$ $R_R = 9/11$ $\boxed{\Phi(\text{split4}|t) = 0.4}$

$$Q(\text{split4}|t) = \left|\overset{7}{\cancel{1/2 - 1/9}}\right| + \left|\overset{5}{\cancel{1/2 - 2/9}}\right| + \left|\overset{8}{\cancel{4/9}}\right| + \left|\overset{4}{\cancel{2/9}}\right| = \boxed{24/18}$$

براساس max Q(split|t) داریم :



برای شاخه چپ :

متغیر age :

split 1 , {low} {mid, hig}

split 2 , {mid} {low, high}

split 3 , {high} {low, mid}

$$\left.\begin{array}{l} \text{split } 1 , \quad P_L = P_R = 1/2 \\ Q(\text{split}1|t) = |1| + |1| = 2 \end{array}\right\} \Rightarrow \phi(\text{split}1|t)$$

$$= 2(1/2)(1/2)(2)$$

$$= 1$$

max $\phi$ است $(k/2)$

برای شاخه راست :

متغیر age :

Split 1 : $P_L = 2/7 \quad P_R = 5/7$

$Q(\text{split}1|t) = |2/2| + |3/5| + |2/5| = 2$

$\Rightarrow \phi(\text{split}1|t) = \boxed{0.8}$

Subject :

Date :

Page :

Split 2 : $P_L = 2/7 \quad P_R = 5/7$

$$Q(\text{split } 2 \mid t) = \left|\overset{0.4}{2/5}\right| + \left|\cancel{2/2 - 1/5}^{\,0.8}\right| + \left|\overset{0.4}{2/5}\right| = \boxed{0.65}$$

Split 3 : $P_L = 3/7 \quad P_R = 4/7$

$$Q(\text{split } 3 \mid t) = \left|\overset{6/12}{\cancel{2/4}}\right| + \left|\cancel{1/3 - 2/4}^{\,2/12}\right| + \left|\overset{8/12}{\cancel{2/3}}\right| = \boxed{0.65}$$

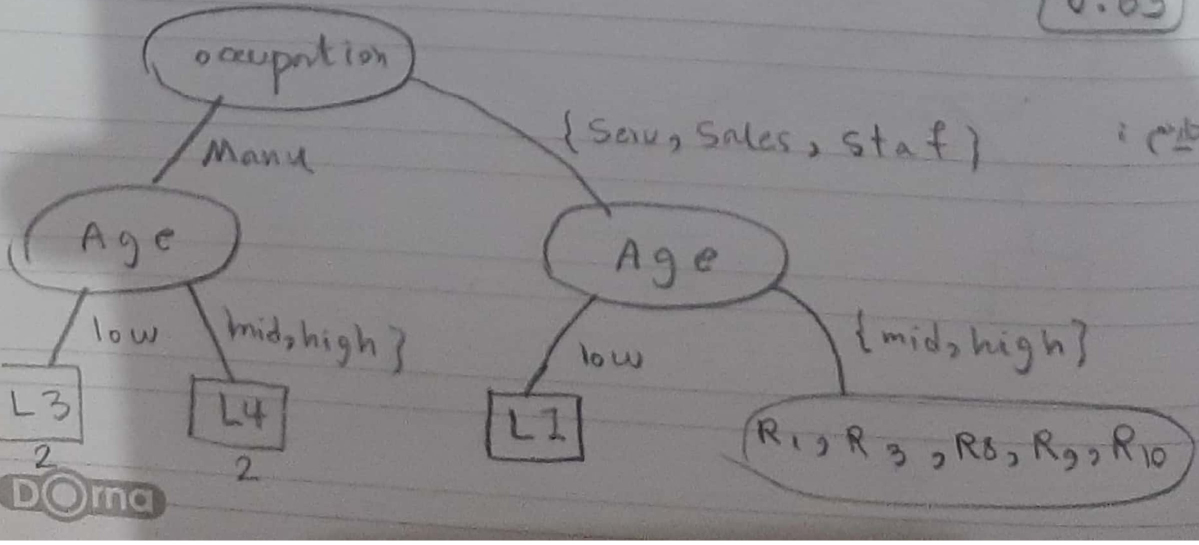
نشر : gender

Split 1 : {Male} {Female}

$P_R = 3/7 \quad P_L = 4/7$

$$Q(\text{split } 1 \mid t) = \left|\overset{6/12}{\cancel{2/4}}\right| + \left|\cancel{2/4 - 1/3}^{\,2/12}\right| + \left|\overset{8/12}{\cancel{2/3}}\right| = \boxed{0.65}$$

طابع :



برای نود پانزده داریم :

split 1 = {Age : mid} {Age : high}

split 2 = {Gender : male} {Gender : female}

split 1 : $P_L = 2/5 \quad P_R = 3/5$

$Q(\text{split}1|t) = \left|\overset{2/3}{\cancel{\frac{2}{2} - \frac{1}{3}}}\right| + \left|\frac{2}{3}\right| = \frac{4}{3}$

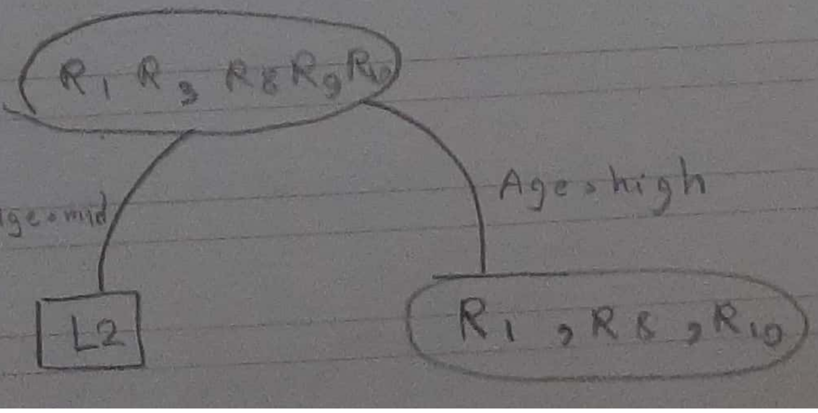
$\Rightarrow \Phi(\text{split }1|t) = 0.64$

split 2 : $P_L = 2/5 \quad P_R = 3/5$

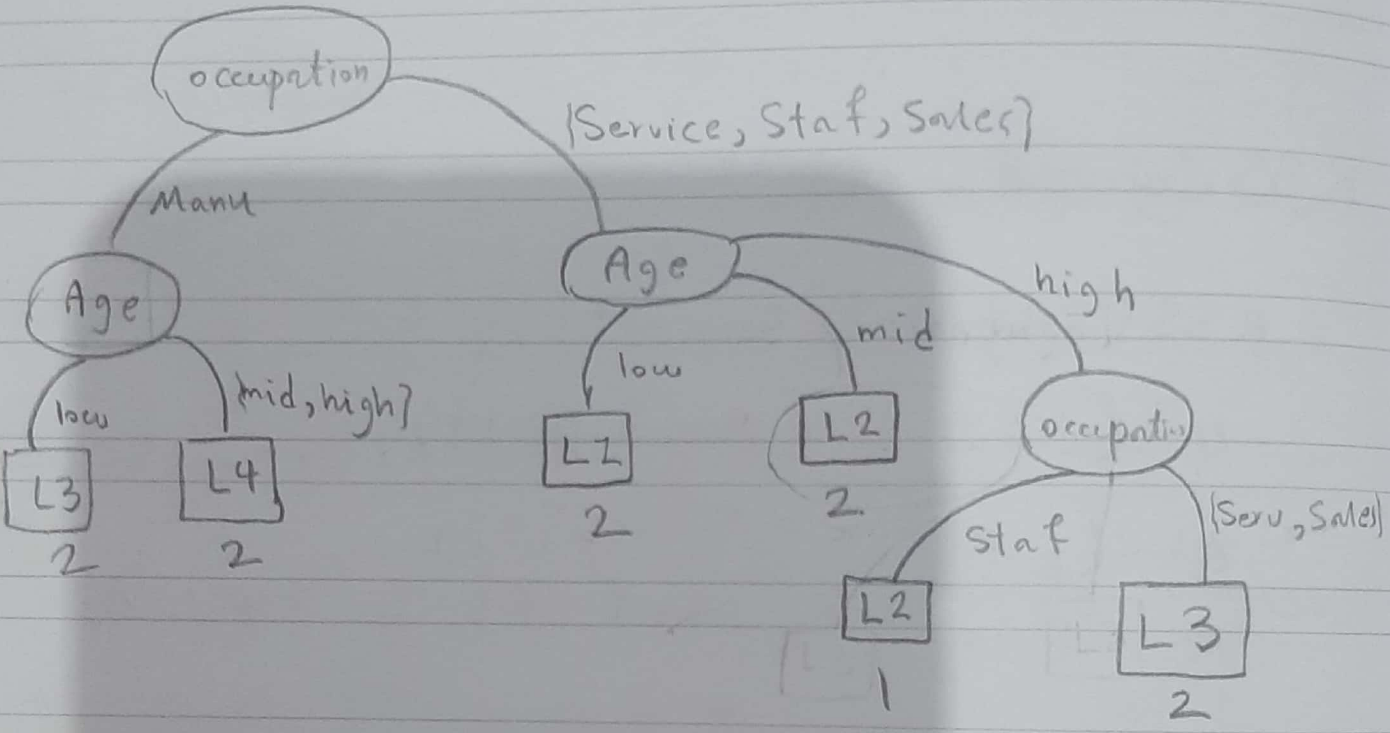
$Q(\text{split}2|t) = \left|\frac{2}{2} - \frac{1}{3}\right| + \left|\frac{2}{3}\right|$

$\Rightarrow \Phi(\text{split }2|t) = 0.64$

از بین 2 split یکی را انتخاب :



3 متغیر R$_1$ ، R$_8$ و R$_{10}$ تنها در یک متغیر فرق دارند و آن هم Ocupation است. پس در نهایت داریم :



همانطور که دیده می‌شود، این درخت (CART) از درخت قبلی (C45) برگ‌ها کمتری دارد (اگر چه عمقش بیشتر است)، این یعنی over fitting کمتری نسبت به درخت قبلی دارد. (دقت شود که در C45 در هر برگ یک sample می‌افتاد و support قوی نداشت، اما در اینجا در برخی برگها بیش از یک sample از training set ما قرار دارند.)

5) ب-

درخت رگرسیون برای متغیر Salary:

RSS (ریشه |t) = 2004.55

متغیر occupation :

RSS (occupation |t)

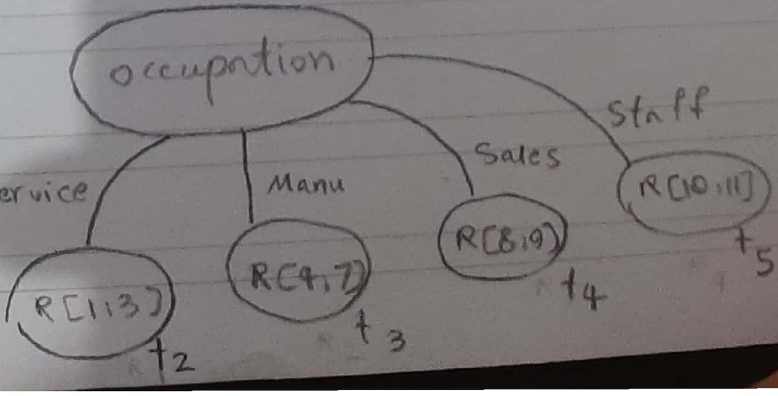
= RSS(occupation = Service |t) +
" ( " = Manufacture |t) +
" ( " = Sales |t) +
" ( " = Staf |t) = 947.25

برای بدست آوردن RSS، ابتدا میانگین Salary های داخل هر زیرگره را حساب کرده و سپس variance آنرا زیرگره را حساب می کنیم

متغیر age : RSS (Age |t) = 1404.67

متغیر Gender : RSS (Gender |t) = 10.44.03

براساس min RSS (یا max Information Gain):

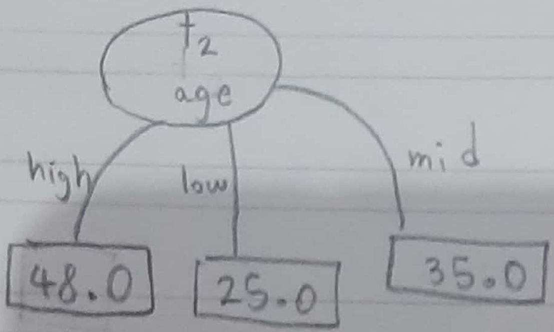


برای شاخه ی (occupation = service):

$RSS(gender \mid t_2) = 50$

$RSS(Age \mid t_2) = 0$

براساس min RSS داریم،

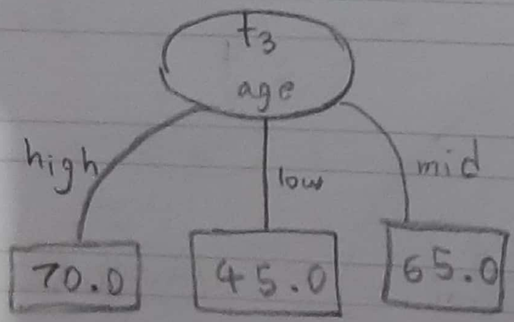


برای شاخه $t_3$ داریم،

$RSS(gender \mid t3) = 12.5$

$RSS(age \mid t_3) = 0$

براساس min RSS داریم،

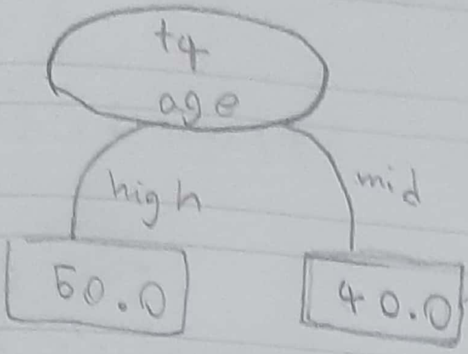


برای شاخه $t_4$ داریم:

$RSS(Age \mid t_4) = 0$

$RSS(Gender \mid t_4) = 0$

حال داریم:

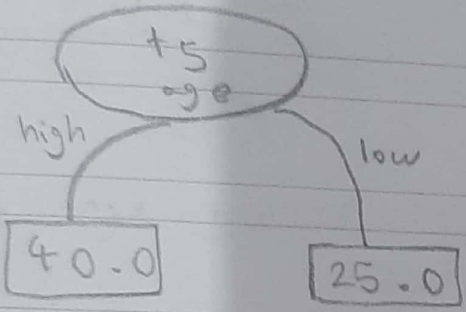


برای شاخه $t_5$ هم داریم:

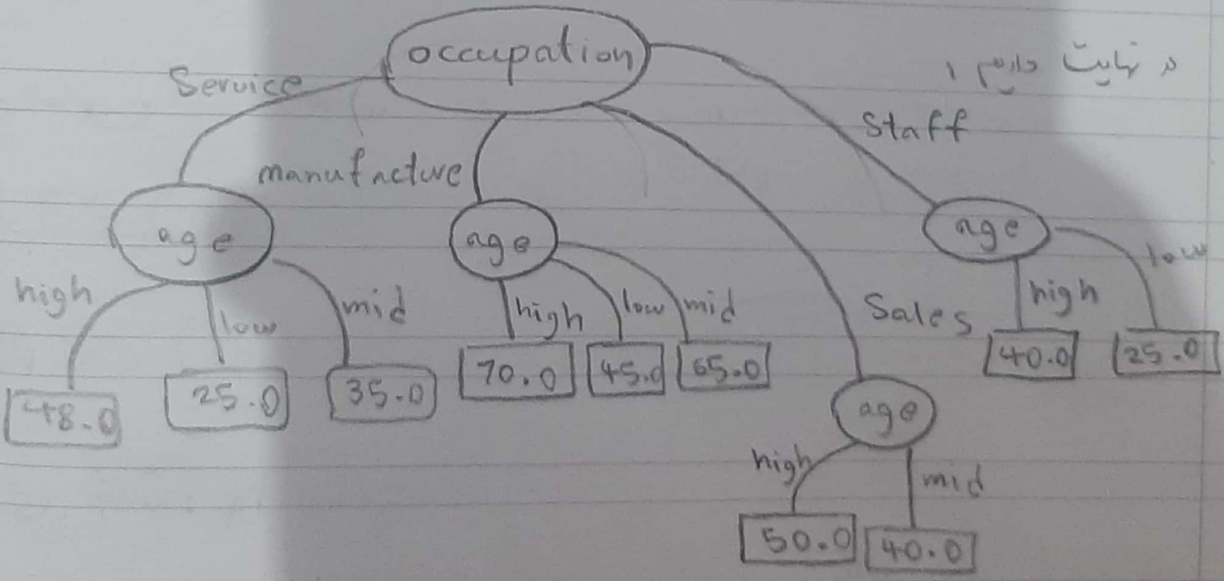
$RSS(Gender \mid t_5) = 0$

$RSS(Age \mid t_5) = 0$

حال داریم:



در نهایت داریم:



همانطور که دیده میشود، چون در صورت سوال شرط خاتمه‌ای ذکر نشده و درخت رگرسیون تا آخرین مرحله پیش رفته، تعداد برگ‌ها زیاد و [illegible] overfitting رخ داده.

7) در درخت بدست آمده از الگوریتم C4.5، تعداد برگ ها بیشتر و درخت bushier است. بنابراین از درخت CART در داده های تست دقت کمتر خواهد بود. (برگ های CART شامل تعداد بیشتری sample هستند، در نتیجه support شان بیشتر است.)

8)

Decision rules of CART tree:

1) if occupation = Manufacture & Age = low
then salary is Level 3 confidence: 1.0
support: 2/11

2) if occupation = Manufacture & Age ∈ {mid, high}
then salary is Level 4 confidence: 1.0
support: 2/11.

3) if occup ∈ {Sales, Staf, Service} & Age = low
then salary is L1 conf: 1.0
support: 2/11

if

4) if occup ∈ {Sales, Staf, Service} & Age = mid then salary is Level 2

sup = 2/11
conf = 1.0

5) if occup = Staf & Age = high then salary is L2

sup = 1.0
conf = 1/11

6) if occup ∈ {Service, Sales} & Age = high then salary is L3

sup = 1.0
conf = 2/11

: C4.5 $\overset{CS}{6}$ Decision rule (9

) If Age = high & occu = Service, then Salary is L3
sup = 1.0
conf = 1/11

2) " " " occup = Manufacture, " " " L4
sup = 1.0
conf = 1/11

3) " " " occup = Sales, " " L3
sup = 1.0
conf = 1/11

4) " " " occup = Staf, " " L2
sup = 1.0
conf = 1/11

5) if age = mid & occup = service, then salary is L2

sup = 1.0

conf = 1/11

6) " " " occup = Manufacture, " " L4

sup = 1.0

conf = 1/11

7) " " " occup = Sales, " " L2

sup = 1.0

conf = 1/11

8) if age = low & occup = Service, then salary is L1

sup = 1.0

conf = 1/11

9) if " " " = Manfacture, " " L3

sup = 1.0

conf = 2/11

10) " " " = Staff, " " L1

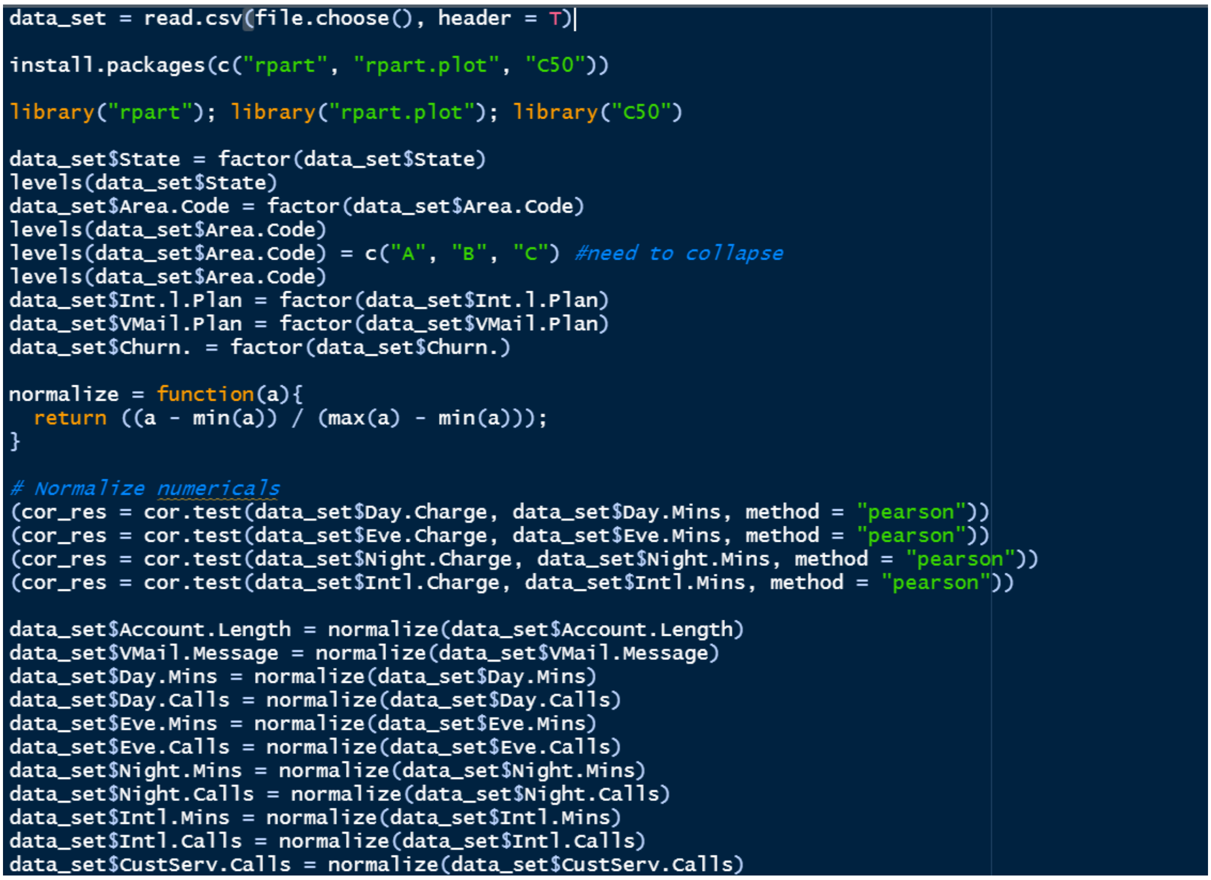
sup = 1.0

conf = 1/11

10) همانطور که دیده میشود، تعداد تراکنش تقسیم c4.5 بیشتر است و support شان هم کمتر از CART می باشد، به عبارتی مدل درخت تقسیم ما در این مثال با c4.5 دچار overfitting شده.

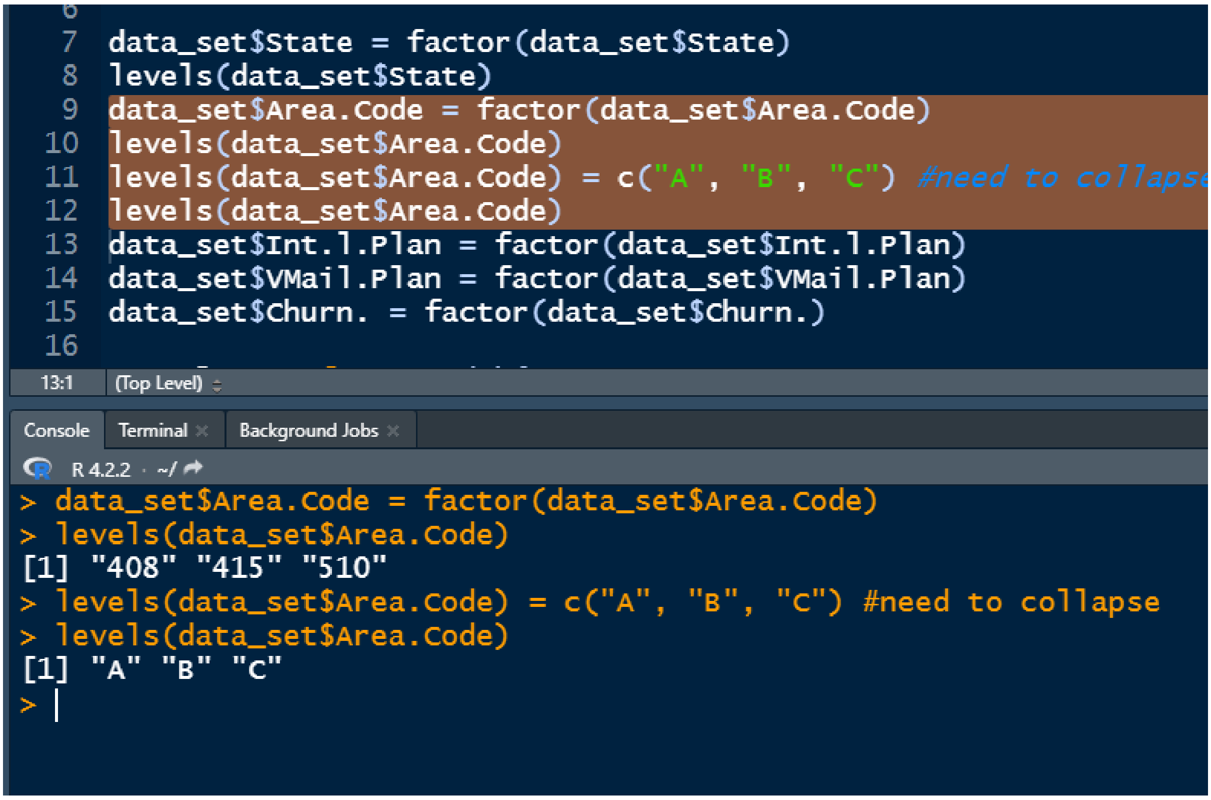
امیرمهدی انصاری پور شماره دانشجویی : 810198358

قبل از انجام الگوریتم های درخت تصمیم، کمی prepocessing انجام میدهیم.



```
data_set = read.csv(file.choose(), header = T)

install.packages(c("rpart", "rpart.plot", "C50"))

library("rpart"); library("rpart.plot"); library("C50")

data_set$State = factor(data_set$State)
levels(data_set$State)
data_set$Area.Code = factor(data_set$Area.Code)
levels(data_set$Area.Code)
levels(data_set$Area.Code) = c("A", "B", "C") #need to collapse
levels(data_set$Area.Code)
data_set$Int.l.Plan = factor(data_set$Int.l.Plan)
data_set$VMail.Plan = factor(data_set$VMail.Plan)
data_set$Churn. = factor(data_set$Churn.)

normalize = function(a){
  return ((a - min(a)) / (max(a) - min(a)));
}

# Normalize numericals
(cor_res = cor.test(data_set$Day.Charge, data_set$Day.Mins, method = "pearson"))
(cor_res = cor.test(data_set$Eve.Charge, data_set$Eve.Mins, method = "pearson"))
(cor_res = cor.test(data_set$Night.Charge, data_set$Night.Mins, method = "pearson"))
(cor_res = cor.test(data_set$Intl.Charge, data_set$Intl.Mins, method = "pearson"))

data_set$Account.Length = normalize(data_set$Account.Length)
data_set$VMail.Message = normalize(data_set$VMail.Message)
data_set$Day.Mins = normalize(data_set$Day.Mins)
data_set$Day.Calls = normalize(data_set$Day.Calls)
data_set$Eve.Mins = normalize(data_set$Eve.Mins)
data_set$Eve.Calls = normalize(data_set$Eve.Calls)
data_set$Night.Mins = normalize(data_set$Night.Mins)
data_set$Night.Calls = normalize(data_set$Night.Calls)
data_set$Intl.Mins = normalize(data_set$Intl.Mins)
data_set$Intl.Calls = normalize(data_set$Intl.Calls)
data_set$CustServ.Calls = normalize(data_set$CustServ.Calls)
```

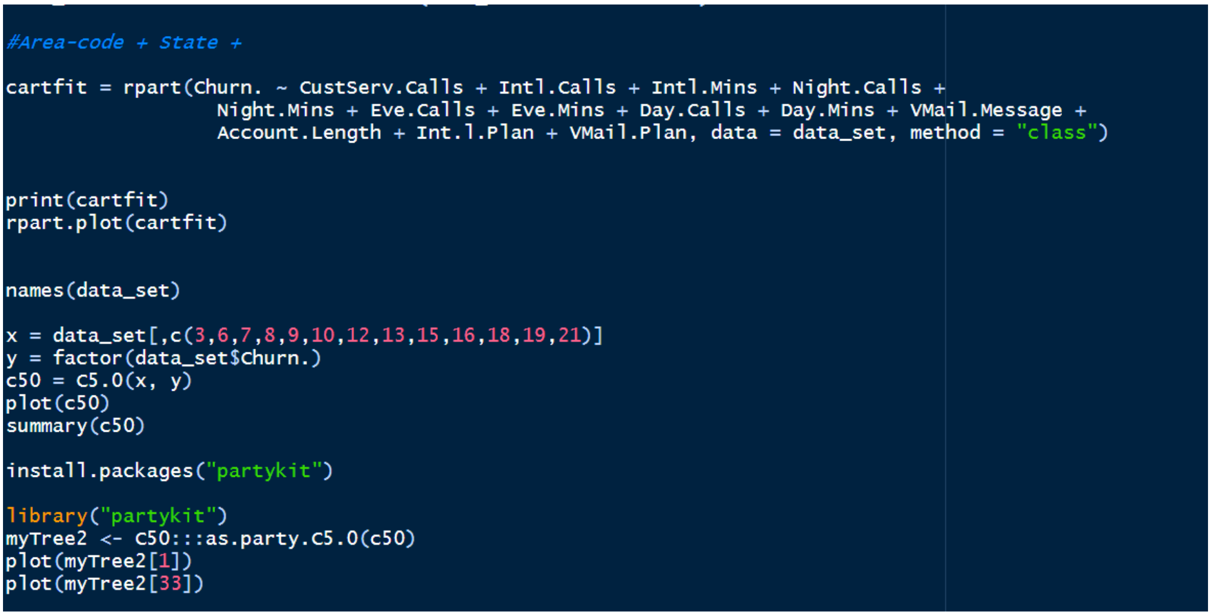
در ابتدا داده های کیفی را بصورت factor درآورده و levelهای آنها را مشاهده میکنیم. داده AreaCode نیاز به collapse دارد زیرا مقادیر آن عددی هستند در حالیکه این متغیر کیفی است و 3 level بیشتر ندارد. (مقادیر 408, 415, 510 که به ترتیب به A,B,C map شده اند)



```
data_set$State = factor(data_set$State)
levels(data_set$State)
data_set$Area.Code = factor(data_set$Area.Code)
levels(data_set$Area.Code)
levels(data_set$Area.Code) = c("A", "B", "C") #need to collaps
levels(data_set$Area.Code)
data_set$Int.l.Plan = factor(data_set$Int.l.Plan)
data_set$VMail.Plan = factor(data_set$VMail.Plan)
data_set$Churn. = factor(data_set$Churn.)
```

```
> data_set$Area.Code = factor(data_set$Area.Code)
> levels(data_set$Area.Code)
[1] "408" "415" "510"
> levels(data_set$Area.Code) = c("A", "B", "C") #need to collapse
> levels(data_set$Area.Code)
[1] "A" "B" "C"
> 
```

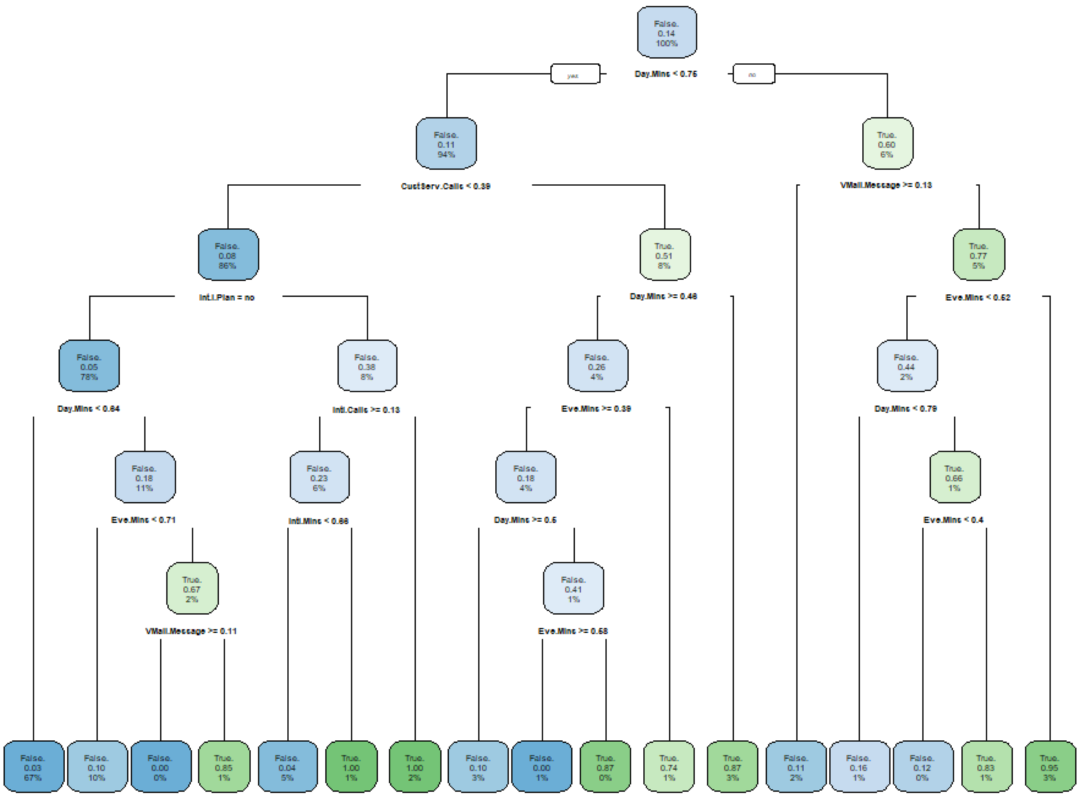
پس از آن داده هایی که با هم correlated هستند را شناسایی کرده و از میان هر جفت correlated attributes، یکی را انتخاب میکنیم. در نهایت نیز داده های کمی را normalize میکنیم.



```
#Area-code + State +

cartfit = rpart(Churn. ~ CustServ.Calls + Intl.Calls + Intl.Mins + Night.Calls +
                  Night.Mins + Eve.Calls + Eve.Mins + Day.Calls + Day.Mins + VMail.Message +
                  Account.Length + Int.l.Plan + VMail.Plan, data = data_set, method = "class")


print(cartfit)
rpart.plot(cartfit)


names(data_set)

x = data_set[,c(3,6,7,8,9,10,12,13,15,16,18,19,21)]
y = factor(data_set$Churn.)
c50 = C5.0(x, y)
plot(c50)
summary(c50)

install.packages("partykit")

library("partykit")
myTree2 <- C50:::as.party.C5.0(c50)
plot(myTree2[1])
plot(myTree2[33])
```

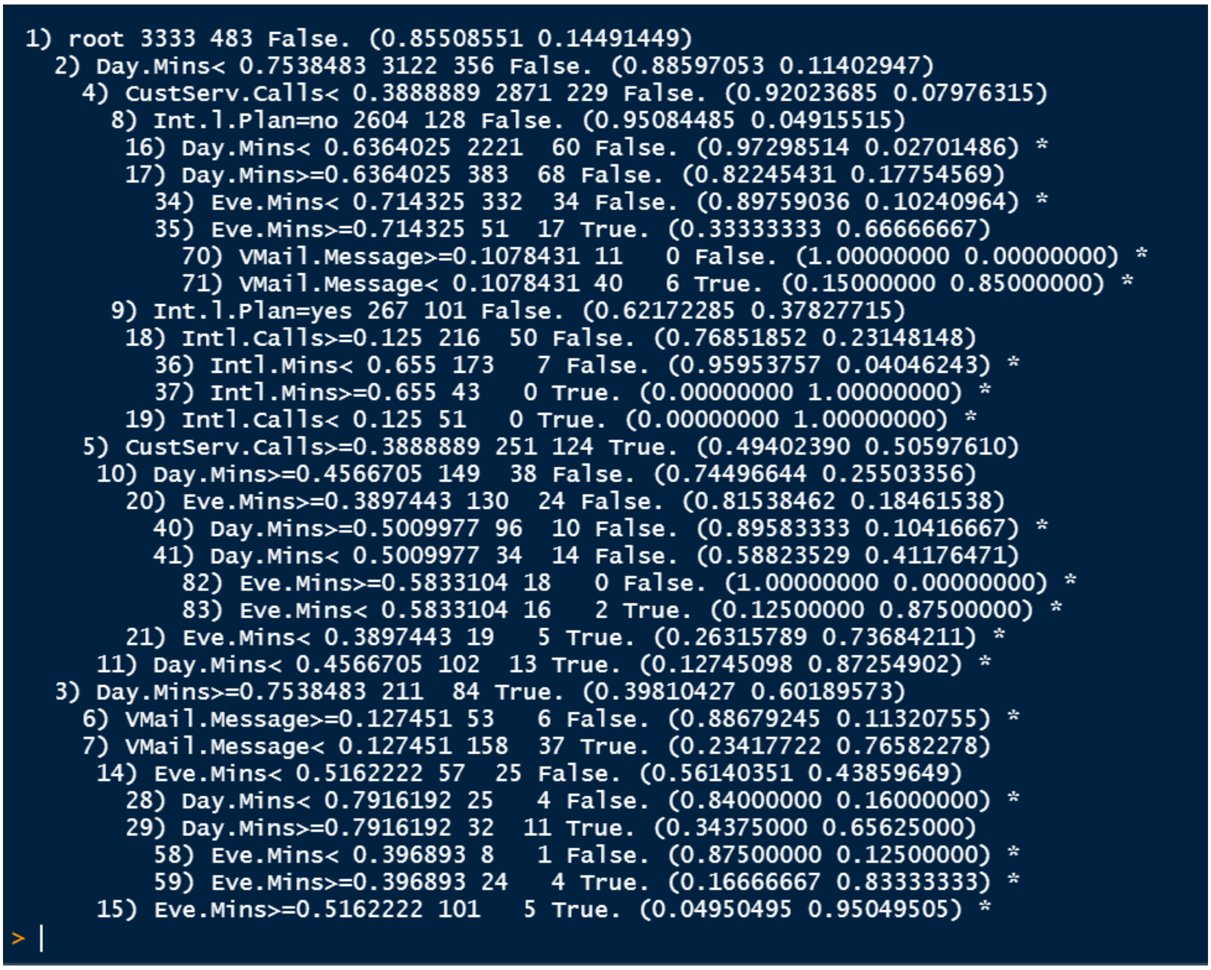
## 11) درخت CART:

با ران کردن کد بالا، خروجی را برای درخت CART میبینیم.



در هر گره به ترتیب کلاس پیشبینی شده، confidence (درصد churn) و support (نسبت تعداد sample های موجود در گره به کل نمونه ها) نمایش داده شده اند.

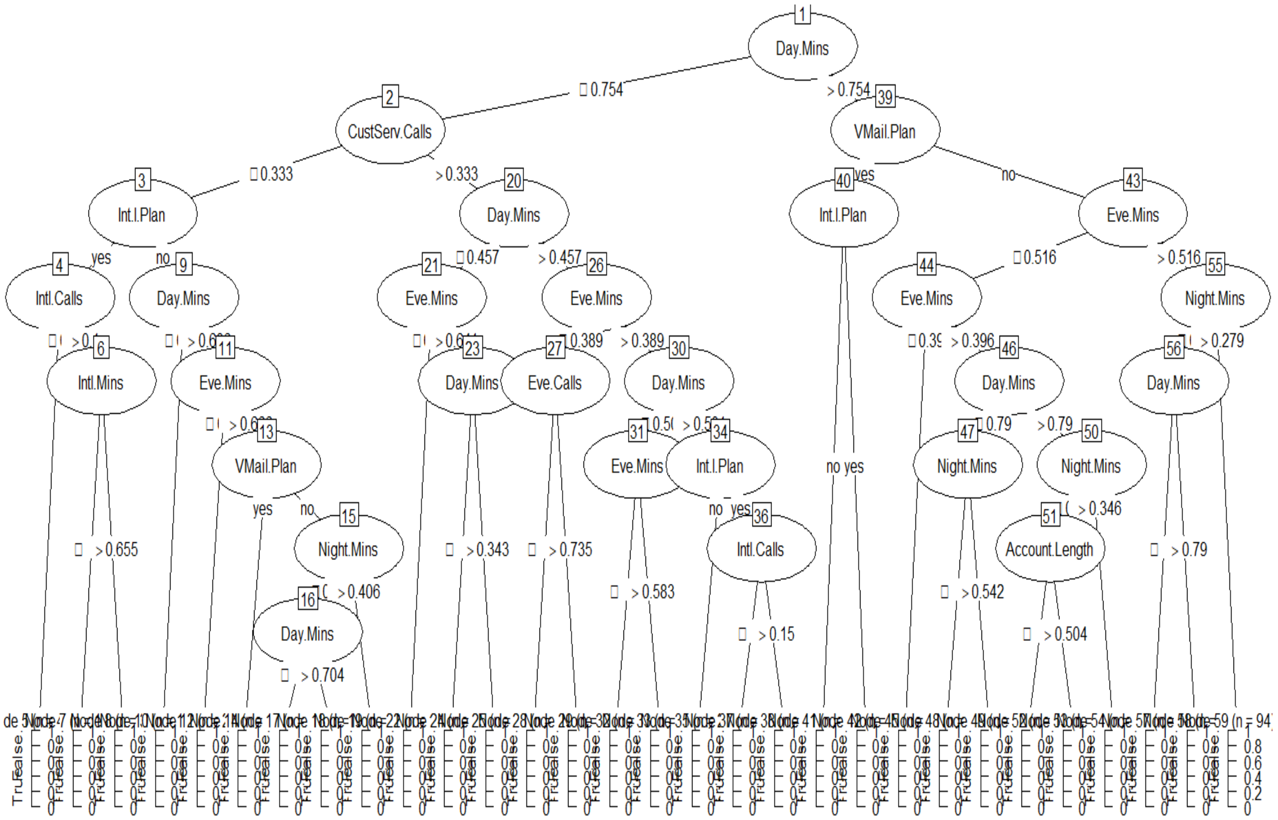
در ادامه decision rules درخت تصمیم نمایش داده شده اند. در هر سطر، ابتدا شرط، سپس تعداد sample ها در زیر درخت چپ و راست، کلاس پیشبینی شده برای آن گره و نسبت هر کدام از آنها به نسبت کل داده های موجود در گره پدر نمایش داده شده است. شرط هایی که به * ختم میشوند، decision rule (برگ) هستند.



```
1) root 3333 483 False. (0.85508551 0.14491449)
  2) Day.Mins< 0.7538483 3122 356 False. (0.88597053 0.11402947)
    4) CustServ.Calls< 0.3888889 2871 229 False. (0.92023685 0.07976315)
      8) Int.l.Plan=no 2604 128 False. (0.95084485 0.04915515)
       16) Day.Mins< 0.6364025 2221  60 False. (0.97298514 0.02701486) *
       17) Day.Mins>=0.6364025 383  68 False. (0.82245431 0.17754569)
         34) Eve.Mins< 0.714325 332  34 False. (0.89759036 0.10240964) *
         35) Eve.Mins>=0.714325 51  17 True. (0.33333333 0.66666667)
           70) VMail.Message>=0.1078431 11   0 False. (1.00000000 0.00000000) *
           71) VMail.Message< 0.1078431 40   6 True. (0.15000000 0.85000000) *
      9) Int.l.Plan=yes 267 101 False. (0.62172285 0.37827715)
       18) Intl.Calls>=0.125 216  50 False. (0.76851852 0.23148148)
         36) Intl.Mins< 0.655 173   7 False. (0.95953757 0.04046243) *
         37) Intl.Mins>=0.655 43   0 True. (0.00000000 1.00000000) *
       19) Intl.Calls< 0.125 51   0 True. (0.00000000 1.00000000) *
    5) CustServ.Calls>=0.3888889 251 124 True. (0.49402390 0.50597610)
     10) Day.Mins>=0.4566705 149  38 False. (0.74496644 0.25503356)
       20) Eve.Mins>=0.3897443 130  24 False. (0.81538462 0.18461538)
         40) Day.Mins>=0.5009977 96  10 False. (0.89583333 0.10416667) *
         41) Day.Mins< 0.5009977 34  14 False. (0.58823529 0.41176471)
           82) Eve.Mins>=0.5833104 18   0 False. (1.00000000 0.00000000) *
           83) Eve.Mins< 0.5833104 16   2 True. (0.12500000 0.87500000) *
       21) Eve.Mins< 0.3897443 19   5 True. (0.26315789 0.73684211) *
     11) Day.Mins< 0.4566705 102  13 True. (0.12745098 0.87254902) *
  3) Day.Mins>=0.7538483 211  84 True. (0.39810427 0.60189573)
    6) VMail.Message>=0.127451 53   6 False. (0.88679245 0.11320755) *
    7) VMail.Message< 0.127451 158  37 True. (0.23417722 0.76582278)
     14) Eve.Mins< 0.5162222 57  25 False. (0.56140351 0.43859649)
       28) Day.Mins< 0.7916192 25   4 False. (0.84000000 0.16000000) *
       29) Day.Mins>=0.7916192 32  11 True. (0.34375000 0.65625000)
         58) Eve.Mins< 0.396893 8   1 False. (0.87500000 0.12500000) *
         59) Eve.Mins>=0.396893 24   4 True. (0.16666667 0.83333333) *
     15) Eve.Mins>=0.5162222 101   5 True. (0.04950495 0.95049505) *
> |
```

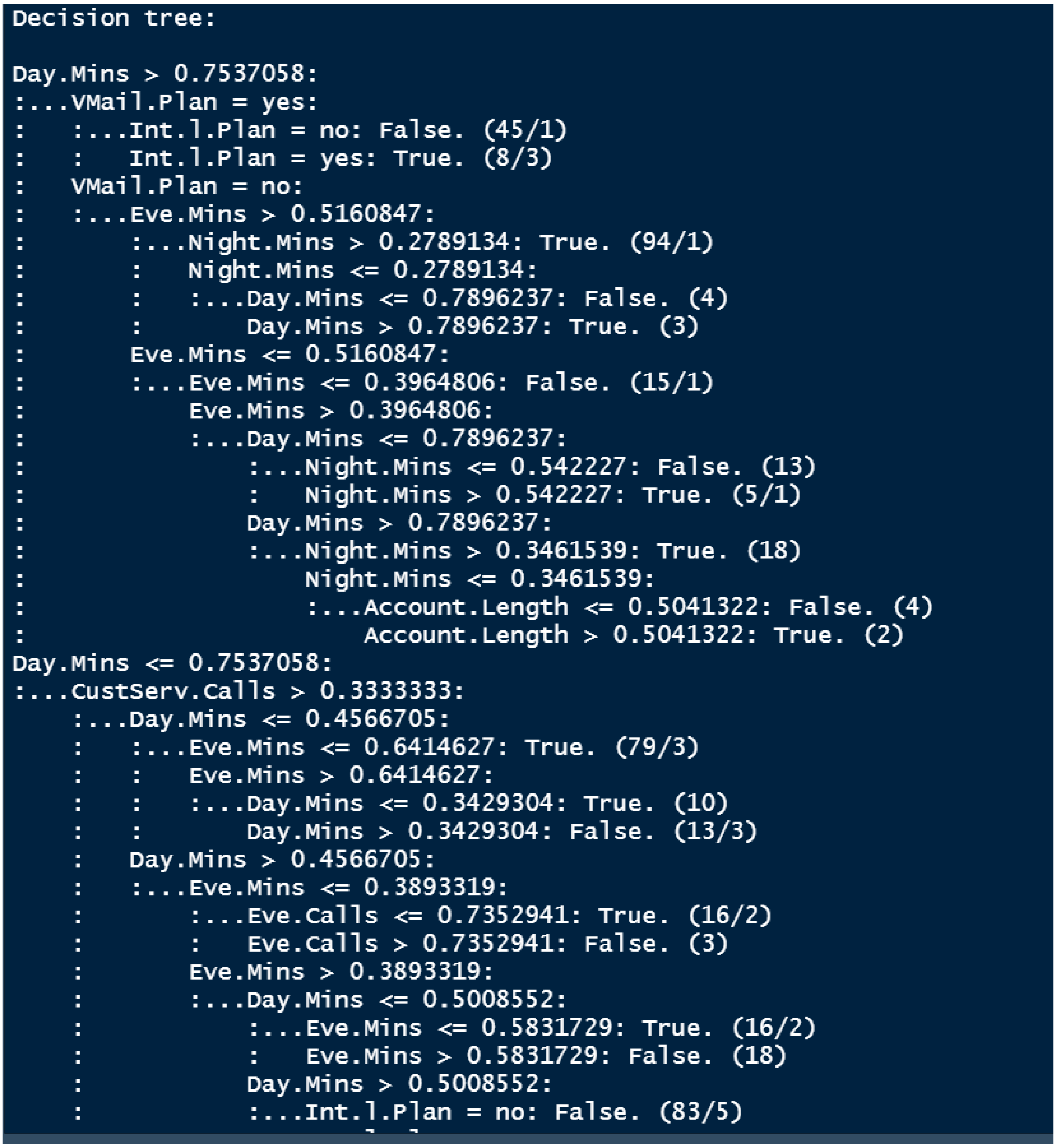
(12

خروجی درخت C4.5:

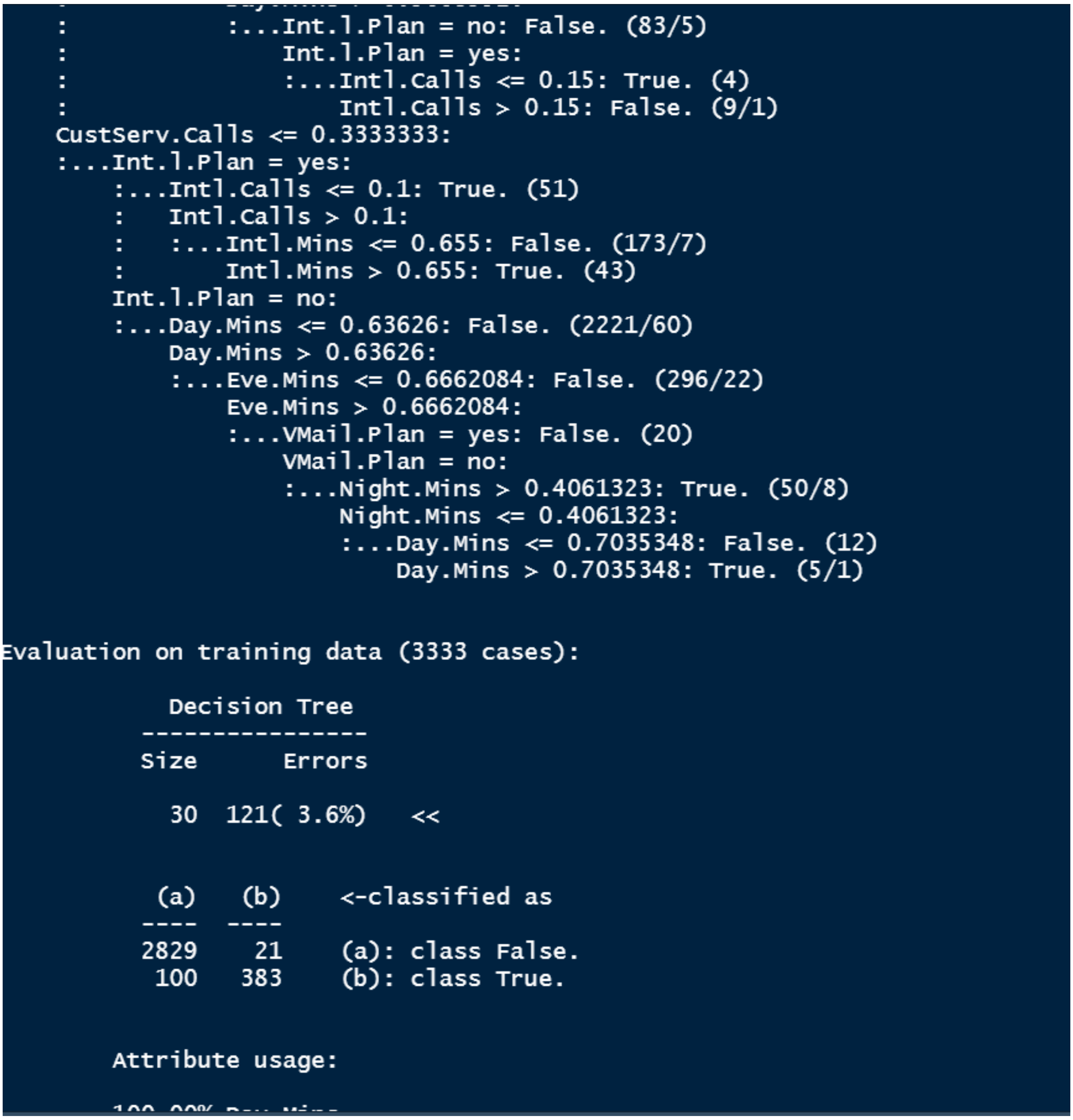


بر روی گره، attribute جداکننده نمونه ها و روی هر یال، نحوه تقسیم بندی نمونه ها قرار دارد.

در ادامه decision rules های درخت تصمیم C4.5 قرار دارد. در هر سطر، ابتدا شرط، سپس کلاس پیش بینی شده برای آن شرط و purity آن گره (تعداد sample های با مقدار متغیر هدف مساوی true و false) نشان داده شده اند. سطرهایی که دیگر ادامه نمی یابند، desicion rule(برگ) هستند.



```
Decision tree:

Day.Mins > 0.7537058:
:...VMail.Plan = yes:
:   :...Int.l.Plan = no: False. (45/1)
:   :   Int.l.Plan = yes: True. (8/3)
:   VMail.Plan = no:
:   :...Eve.Mins > 0.5160847:
:       :...Night.Mins > 0.2789134: True. (94/1)
:       :   Night.Mins <= 0.2789134:
:       :   :...Day.Mins <= 0.7896237: False. (4)
:       :       Day.Mins > 0.7896237: True. (3)
:       Eve.Mins <= 0.5160847:
:       :...Eve.Mins <= 0.3964806: False. (15/1)
:           Eve.Mins > 0.3964806:
:           :...Day.Mins <= 0.7896237:
:               :...Night.Mins <= 0.542227: False. (13)
:               :   Night.Mins > 0.542227: True. (5/1)
:               Day.Mins > 0.7896237:
:               :...Night.Mins > 0.3461539: True. (18)
:                   Night.Mins <= 0.3461539:
:                   :...Account.Length <= 0.5041322: False. (4)
:                       Account.Length > 0.5041322: True. (2)
Day.Mins <= 0.7537058:
:...CustServ.Calls > 0.3333333:
    :...Day.Mins <= 0.4566705:
    :   :...Eve.Mins <= 0.6414627: True. (79/3)
    :   :   Eve.Mins > 0.6414627:
    :   :   :...Day.Mins <= 0.3429304: True. (10)
    :   :       Day.Mins > 0.3429304: False. (13/3)
    :   Day.Mins > 0.4566705:
    :   :...Eve.Mins <= 0.3893319:
    :       :...Eve.Calls <= 0.7352941: True. (16/2)
    :       :   Eve.Calls > 0.7352941: False. (3)
    :       Eve.Mins > 0.3893319:
    :       :...Day.Mins <= 0.5008552:
    :           :...Eve.Mins <= 0.5831729: True. (16/2)
    :           :   Eve.Mins > 0.5831729: False. (18)
    :           Day.Mins > 0.5008552:
    :           :...Int.l.Plan = no: False. (83/5)
```



```
    :           :...Int.l.Plan = no: False. (83/5)
    :               Int.l.Plan = yes:
    :               :...Intl.Calls <= 0.15: True. (4)
    :                   Intl.Calls > 0.15: False. (9/1)
    CustServ.Calls <= 0.3333333:
    :...Int.l.Plan = yes:
        :...Intl.Calls <= 0.1: True. (51)
        :   Intl.Calls > 0.1:
        :   :...Intl.Mins <= 0.655: False. (173/7)
        :       Intl.Mins > 0.655: True. (43)
        Int.l.Plan = no:
        :...Day.Mins <= 0.63626: False. (2221/60)
            Day.Mins > 0.63626:
            :...Eve.Mins <= 0.6662084: False. (296/22)
                Eve.Mins > 0.6662084:
                :...VMail.Plan = yes: False. (20)
                    VMail.Plan = no:
                    :...Night.Mins > 0.4061323: True. (50/8)
                        Night.Mins <= 0.4061323:
                        :...Day.Mins <= 0.7035348: False. (12)
                            Day.Mins > 0.7035348: True. (5/1)


Evaluation on training data (3333 cases):

            Decision Tree
          ----------------
          Size      Errors

            30  121( 3.6%)   <<


           (a)   (b)    <-classified as
          ----  ----
          2829    21    (a): class False.
           100   383    (b): class True.


        Attribute usage:
```

13) همانطور که دیده شد، درخت حاصل از الگوریتم C4.5 دارای برگ های بیشتر و decision rule های بیشتری است. این بدان معناست که پیچیدگی این درخت بیشتر از درخت CART می باشد. پیچیدگی زیاد منجر به overfitting میشود.