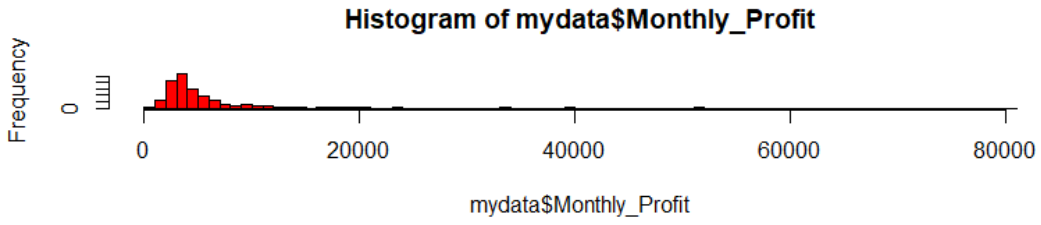
מגישות: 313560484, 205567944

עבודה 2 מדעי הנתונים

1. ביצוע הדיסקרטיזציה:
   - עבור Monthly_Profit השתמשנו ב- Equal-frequency discretization.
     ניתן לראות לפי ההיסטורגמה של התפלגות הערכים שהרוב המוחלט של הערכים קטן מ-20,000 ועל כן רצינו ליצור חלוקה של שווה של ערכים בין ה-bins.
     בחרנו עבור משתנה זה ב-5 bins.



   - עבור Spouse_Income השתמשנו בשיטה Equal-frequency discretization.
     הבחנו כי הערך "0" הופיע בכמות רבה של רשומות (מעל 400 רשומות), על כן החלטנו לחלק את הרשומות באופן שבו בכל אינטרוול מספר ערכים שווה.
     בחרנו עבור משתנה זה ב-5 bins.

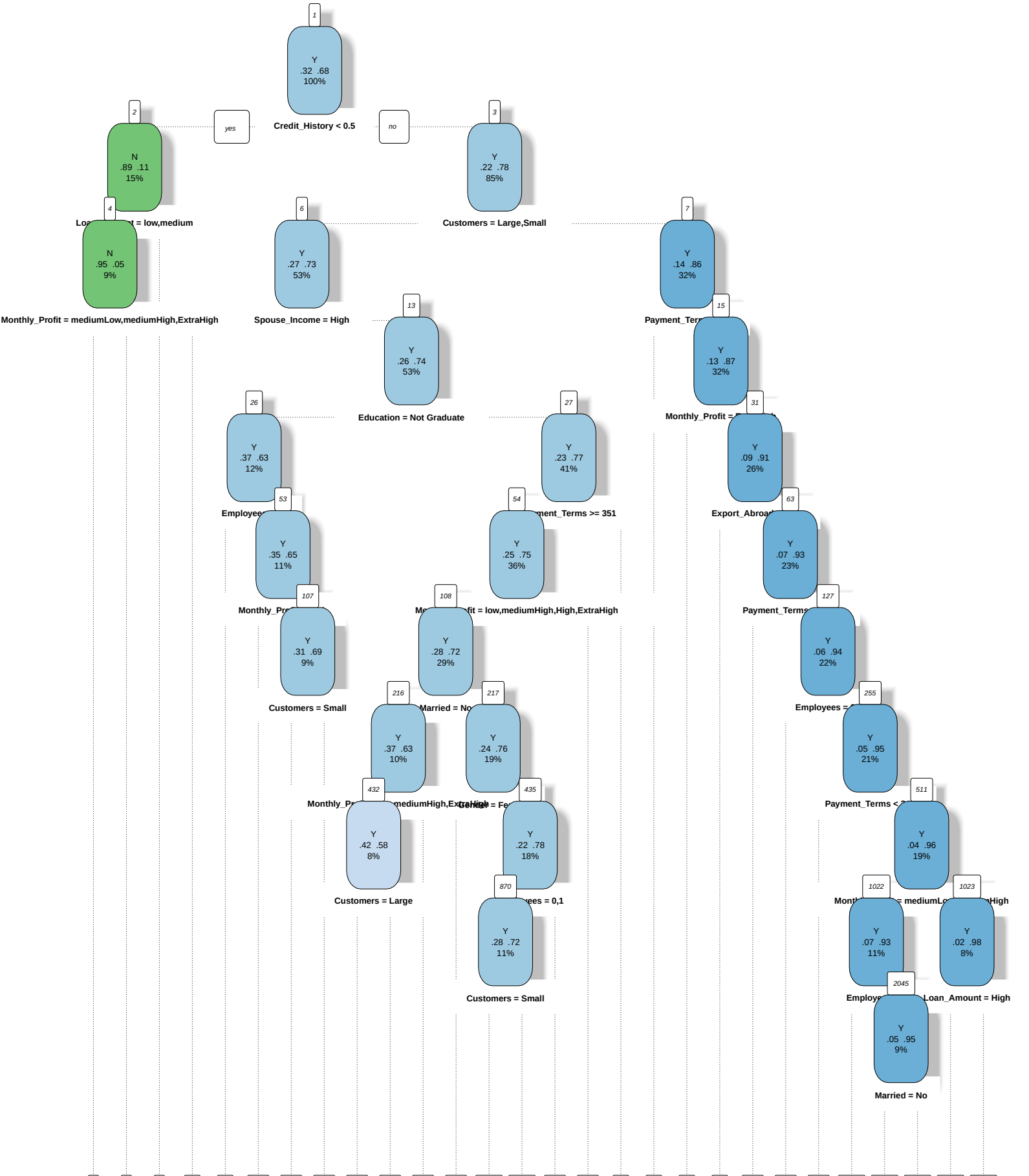
   - עבור Loan_Amount השתמשנו בשיטה Equal-frequency discretization.
     טווח הערכים במשתנה זה קטן יותר ולכן בחרנו לחלק משתנה זה ל-3 bins.

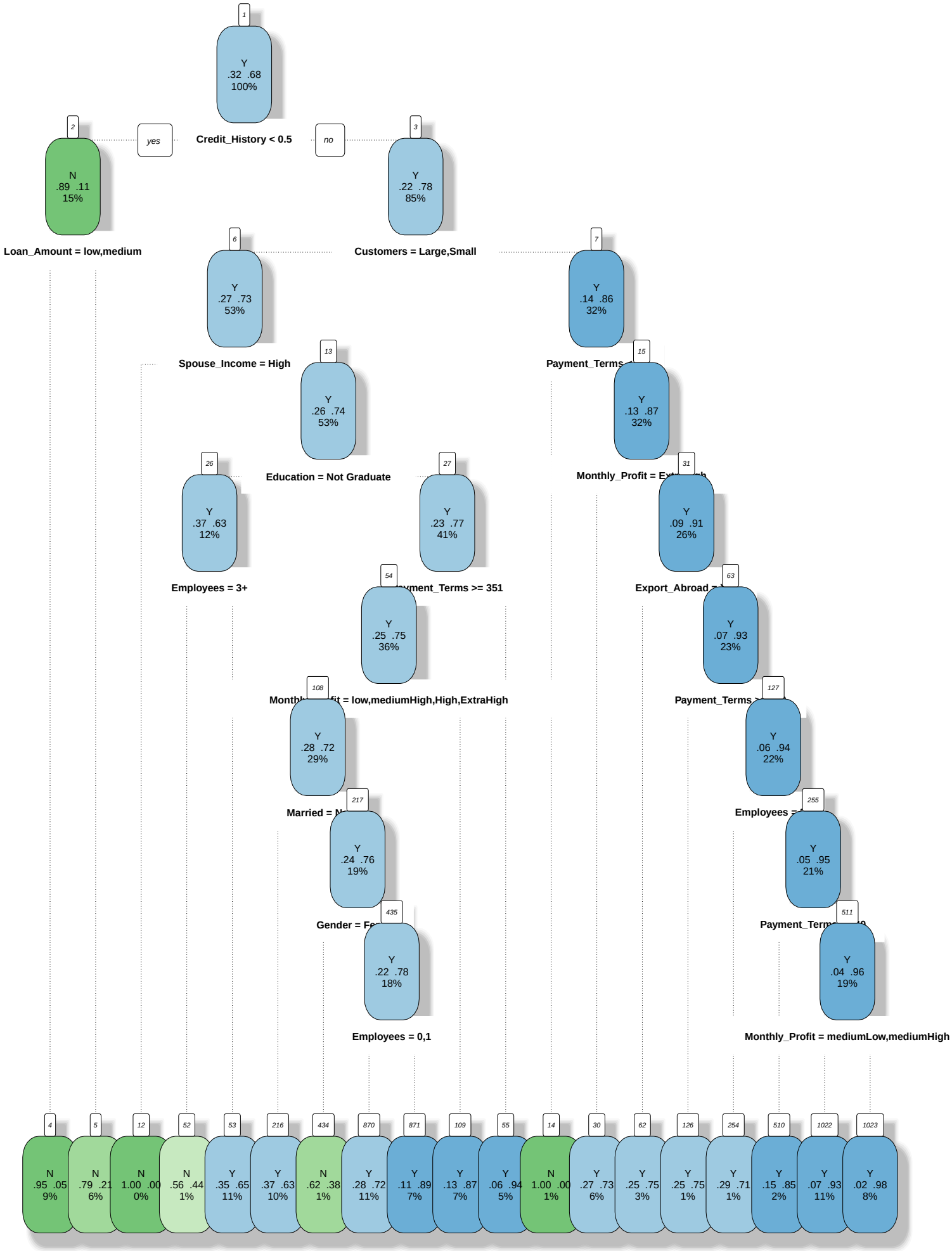
2. תצלומי העצים שהתקבלו מופיעים ברצף בדפים הבאים לפי הסדר הבא:
   - Gini 50
   - Gini 80
   - Information 50
   - Information 80

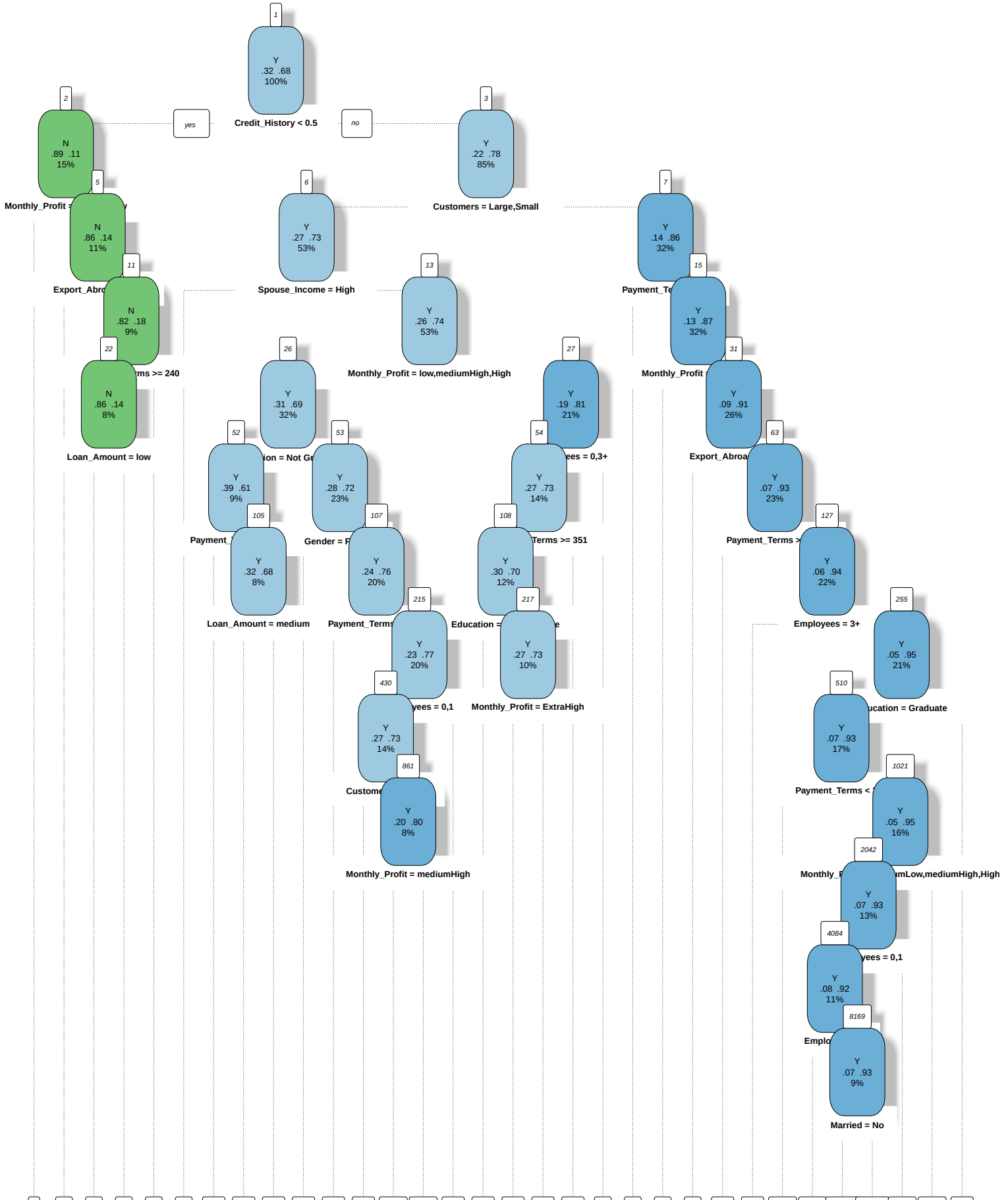
3.

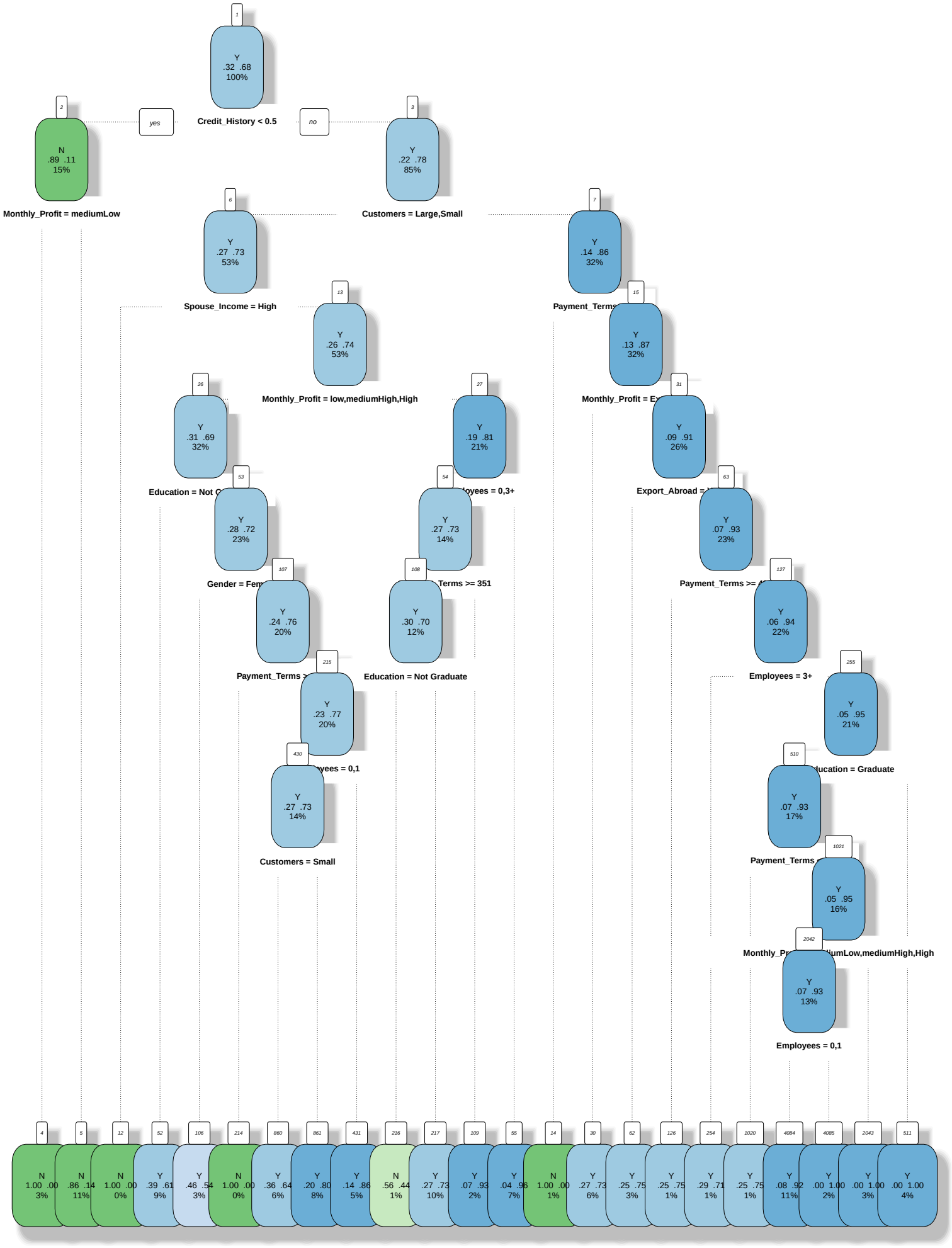
Model	Accuracy
Gini30	0.7726
Information30	0.7893
Gini80	0.796
Information80	0.796

ניתן לראות כי הדיוק הטוב ביותר מתקבל כאשר משתמשים ב-minSplit=80 לא משנה באיזה מודל (gini או information).


1
Y
.32 .68
100%
Credit_History < 0.5
yes
no
2
N
.89 .11
15%
3
Y
.22 .78
85%
4
N
.95 .05
9%
Monthly_Profit = mediumLow,mediumHigh,ExtraHigh
6
Y
.27 .73
53%
Spouse_Income = High
Customers = Large,Small
7
Y
.14 .86
32%
13
Y
.26 .74
53%
Education = Not Graduate
15
Y
.13 .87
32%
26
Y
.37 .63
12%
27
Y
.23 .77
41%
31
Y
.09 .91
26%
53
Y
.35 .65
11%
54
Y
.25 .75
36%
63
Y
.07 .93
23%
107
Y
.31 .69
9%
Customers = Small
108
Y
.28 .72
29%
127
Y
.06 .94
22%
216
Y
.37 .63
10%
217
Y
.24 .76
19%
255
Y
.05 .95
21%
432
Y
.42 .58
8%
Customers = Large
435
Y
.22 .78
18%
511
Y
.04 .96
19%
870
Y
.28 .72
11%
Customers = Small
1022
Y
.07 .93
11%
1023
Y
.02 .98
8%
2045
Y
.05 .95
9%
Married = No
Loan_Amount = High
8 9 5 12 52 106 214 215 864 865 433 434 1740 1741 871 109 55 14 30 62 126 254 510 2044 4090 4091 2046 2047
N 1.00 5%
N .90 .10 4%
N .79 .21 6%
N 1.00 0%
N .56 .44 1%
N .54 .46 2%
Y .37 .63 5%
Y .22 .78 4%
N .44 .56 4%
Y .30 .70 4%
Y .14 .86 2%
N .62 .38 1%
Y .53 .47 5%
Y .14 .86 6%
Y .11 .89 7%
Y .13 .87 7%
Y .06 .94 5%
N 1.00 1%
Y .27 .73 6%
Y .25 .75 3%
Y .25 .75 1%
Y .29 .71 1%
Y .15 .85 2%
Y .17 .83 2%
Y .12 .88 4%
Y .00 1.00 5%
Y .07 .93 2%
Y .00 1.00 6%


1
Y
.32 .68
100%
Credit_History < 0.5
yes
no
2
N
.89 .11
15%
Loan_Amount = low,medium
3
Y
.22 .78
85%
Customers = Large,Small
6
Y
.27 .73
53%
Spouse_Income = High
7
Y
.14 .86
32%
Payment_Terms
13
Y
.26 .74
53%
Education = Not Graduate
15
Y
.13 .87
32%
Monthly_Profit =
26
Y
.37 .63
12%
Employees = 3+
27
Y
.23 .77
41%
ment_Terms >= 351
31
Y
.09 .91
26%
Export_Abroad =
54
Y
.25 .75
36%
fit = low,mediumHigh,High,ExtraHigh
63
Y
.07 .93
23%
Payment_Terms
108
Y
.28 .72
29%
Married =
127
Y
.06 .94
22%
Employees =
217
Y
.24 .76
19%
Gender =
255
Y
.05 .95
21%
Payment_Term
435
Y
.22 .78
18%
Employees = 0,1
511
Y
.04 .96
19%
Monthly_Profit = mediumLow,mediumHigh
4
N
.95 .05
9%
5
N
.79 .21
6%
12
N
1.00 .00
0%
52
N
.56 .44
1%
53
Y
.35 .65
11%
216
Y
.37 .63
10%
434
N
.62 .38
1%
870
Y
.28 .72
11%
871
Y
.11 .89
7%
109
Y
.13 .87
7%
55
Y
.06 .94
5%
14
N
1.00 .00
1%
30
Y
.27 .73
6%
62
Y
.25 .75
3%
126
Y
.25 .75
1%
254
Y
.29 .71
1%
510
Y
.15 .85
2%
1022
Y
.07 .93
11%
1023
Y
.02 .98
8%


1
Y .32 .68 100%
Credit_History < 0.5
yes
no
2
N .89 .11 15%
Monthly_Profit
3
Y .22 .78 85%
Customers = Large,Small
5
N .86 .14 11%
Export_Abro
6
Y .27 .73 53%
Spouse_Income = High
7
Y .14 .86 32%
Payment_Te
11
N .82 .18 9%
rms >= 240
13
Y .26 .74 53%
Monthly_Profit = low,mediumHigh,High
15
Y .13 .87 32%
Monthly_Profit
22
N .86 .14 8%
Loan_Amount = low
26
Y .31 .69 32%
ion = Not Gr
27
Y .19 .81 21%
ees = 0,3+
31
Y .09 .91 26%
Export_Abroa
52
Y .39 .61 9%
Payment_
53
Y .28 .72 23%
Gender = F
54
Y .27 .73 14%
Terms >= 351
63
Y .07 .93 23%
Payment_Terms >
105
Y .32 .68 8%
Loan_Amount = medium
107
Y .24 .76 20%
Payment_Terms
108
Y .30 .70 12%
Education =
127
Y .06 .94 22%
Employees = 3+
215
Y .23 .77 20%
yees = 0,1
217
Y .27 .73 10%
Monthly_Profit = ExtraHigh
255
Y .05 .95 21%
ucation = Graduate
430
Y .27 .73 14%
Custome
510
Y .07 .93 17%
Payment_Terms <
861
Y .20 .80 8%
Monthly_Profit = mediumHigh
1021
Y .05 .95 16%
Monthly_
mLow,mediumHigh,High
2042
Y .07 .93 13%
yees = 0,1
4084
Y .08 .92 11%
Emplo
8169
Y .07 .93 9%
Married = No
4 10 44 45 23 12 104 210 211 106 214 860 1722 1723 431 216 434 435 109 55 14 30 62 126 254 1020 8168 16338 16339 4085 2043 511
N 1.00 .00 3%
N 1.00 .00 2%
N 1.00 .00 3%
N .00 79 6%
Y .21 .25 .75 1%
N 1.00 .00 0%
N .87 .12 1%
Y .46 .54 4%
Y .19 .81 4%
Y .46 .54 3%
N 1.00 .00 0%
Y .36 .64 6%
N .50 .50 1%
Y .13 .87 6%
Y .14 .86 5%
N .86 .56 1%
Y .44 .37 5%
Y .63 .17 5%
Y .83 .07 2%
Y .93 .04 7%
N .96 1.00 1%
Y .00 27 6%
Y .73 .25 3%
Y .75 .25 1%
Y .75 .29 1%
Y .71 .25 1%
Y .75 .14 2%
Y .86 .11 4%
Y .89 .03 5%
Y .97 .00 2%
Y 1.00 .00 3%
Y 1.00 .00 4%


1
Y
.32 .68
100%
Credit_History < 0.5
yes
no
2
N
.89 .11
15%
Monthly_Profit = mediumLow
3
Y
.22 .78
85%
Customers = Large,Small
6
Y
.27 .73
53%
Spouse_Income = High
7
Y
.14 .86
32%
Payment_Terms
13
Y
.26 .74
53%
Monthly_Profit = low,mediumHigh,High
15
Y
.13 .87
32%
Monthly_Profit = E
26
Y
.31 .69
32%
Education = Not G
27
Y
.19 .81
21%
oyees = 0,3+
31
Y
.09 .91
26%
Export_Abroad =
53
Y
.28 .72
23%
Gender = Fem
54
Y
.27 .73
14%
Terms >= 351
63
Y
.07 .93
23%
Payment_Terms >= 4
107
Y
.24 .76
20%
Payment_Terms >
108
Y
.30 .70
12%
Education = Not Graduate
127
Y
.06 .94
22%
Employees = 3+
215
Y
.23 .77
20%
yees = 0,1
255
Y
.05 .95
21%
ucation = Graduate
430
Y
.27 .73
14%
Customers = Small
510
Y
.07 .93
17%
Payment_Terms
1021
Y
.05 .95
16%
Monthly_Pr
iumLow,mediumHigh,High
2042
Y
.07 .93
13%
Employees = 0,1
4
N
1.00 .00
3%
5
N
.86 .14
11%
12
N
1.00 .00
0%
52
Y
.39 .61
9%
106
Y
.46 .54
3%
214
N
1.00 .00
0%
860
Y
.36 .64
6%
861
Y
.20 .80
8%
431
Y
.14 .86
5%
216
N
.56 .44
1%
217
Y
.27 .73
10%
109
Y
.07 .93
2%
55
Y
.04 .96
7%
14
N
1.00 .00
1%
30
Y
.27 .73
6%
62
Y
.25 .75
3%
126
Y
.25 .75
1%
254
Y
.29 .71
1%
1020
Y
.25 .75
1%
4084
Y
.08 .92
11%
4085
Y
.00 1.00
2%
2043
Y
.00 1.00
3%
511
Y
.00 1.00
4%