

Data Warehousing and Data Mining

Prashanth.S (19MID0020)

Question-2 Decision Tree

.arff file

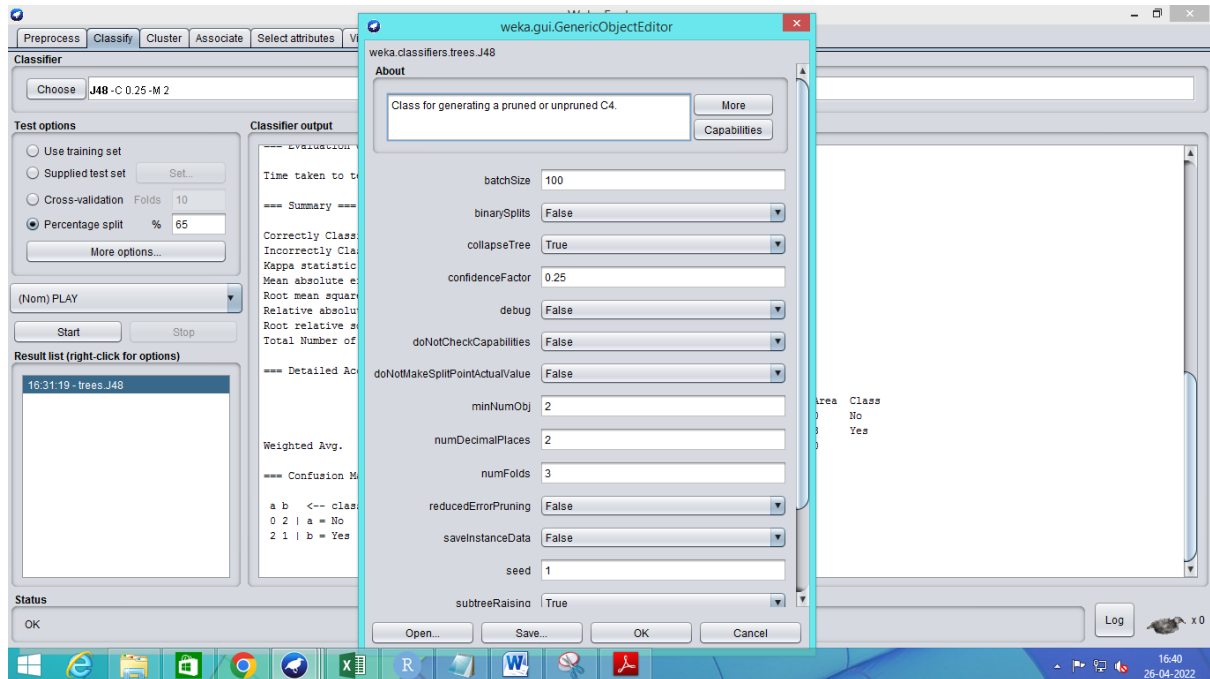
```
data - Notepad
File Edit Format View Help
@RELATION TABLE
@ATTRIBUTE DAY NUMERIC
@ATTRIBUTE OUTLOOK {Sunny,Overcast,Rain}
@ATTRIBUTE HUMIDITY {High,Normal}
@ATTRIBUTE WIND {Weak,Strong}
@ATTRIBUTE PLAY {No,Yes}
@DATA
1,Sunny,High,Weak,No
2,Sunny,High,Strong,No
3,Overcast,High,Weak,Yes
4,Rain,High,Weak,Yes
5,Rain,Normal,Weak,Yes
6,Rain,Normal,Strong,No
7,Overcast,Normal,Strong,Yes
8,Sunny,High,Weak,No
9,Sunny,Normal,Weak,Yes
10,Rain,Normal,Weak,Yes
11,Sunny,Normal,Strong,Yes
12,Overcast,High,Strong,Yes
13,Overcast,Normal,Weak,Yes
14,Rain,High,Strong,No
15,Rain,High,Weak,
```

Successfully imported the data-set

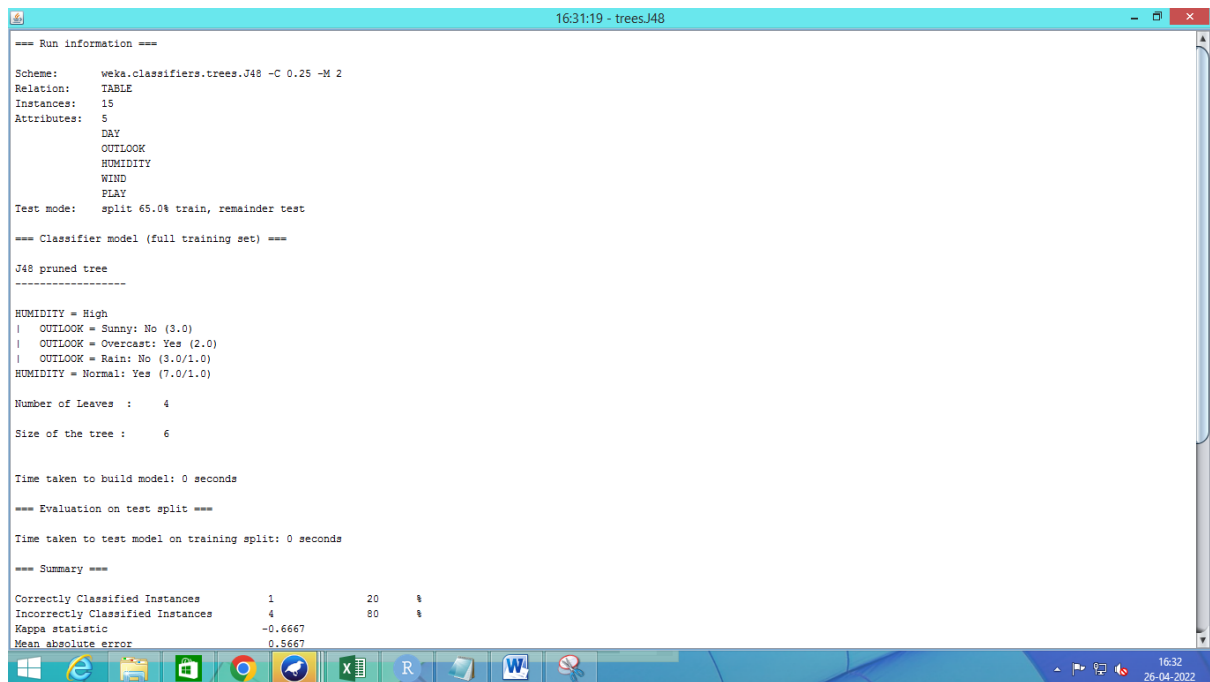
The screenshot shows the Weka Explorer interface. On the left, a 'Viewer' window displays the dataset details, including a table of 15 instances with attributes: DAY (Numeric), OUTLOOK (Nominal), HUMIDITY (Nominal), WIND (Nominal), and PLAY (Nominal). The 'PLAY' attribute is selected, and its statistics are shown: Name: PLAY, Missing: 0 (0%), Distinct: 2, Type: Nominal, Unique: 0 (0%). Below this, a table shows the distribution of the 'PLAY' attribute: 'No' (6 instances, weight 6.0) and 'Yes' (9 instances, weight 9.0). At the bottom, a visualization shows two colored rectangles: a blue rectangle labeled '0' and a red rectangle labeled '9', representing the distribution of the 'PLAY' attribute.

No.	Label	Count	Weight
1	No	6	6.0
2	Yes	9	9.0

Decision Tree



Weka Output



```
16:31:19 - trees.J48

| OUTLOOK = Sunny: No (3.0)
| OUTLOOK = Overcast: Yes (2.0)
| OUTLOOK = Rain: No (3.0/1.0)
HUMIDITY = Normal: Yes (7.0/1.0)

Number of Leaves :    4
Size of the tree :    6

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0 seconds

=== Summary ===

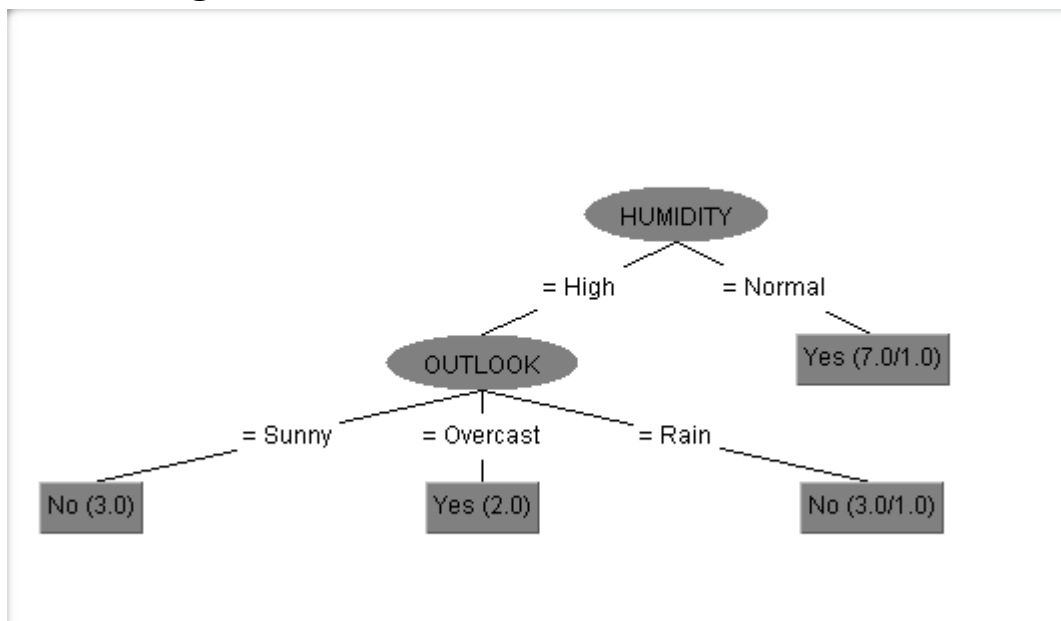
Correctly Classified Instances      1      20  %
Incorrectly Classified Instances    4      80  %
Kappa statistic                    -0.6667
Mean absolute error                 0.5667
Root mean squared error             0.6346
Relative absolute error             117.2414 %
Root relative squared error         129.472 %
Total Number of Instances          5

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.000   0.667   0.000   0.000   0.000   -0.667   0.333   0.500   No
0.333   1.000   0.333   0.333   0.333   -0.667   0.333   0.733   Yes
Weighted Avg.   0.200   0.867   0.200   0.200   0.200   -0.667   0.333   0.640

=== Confusion Matrix ===
a b  <-- classified as
0 2 | a = No
2 1 | b = Yes
```

Visualizing the tree



Output:

From the above tree it is clear that

Outlook : Rain

Humidity : High

Wind: Weak

Ans: The player won't go for the game

QUESTION-1

DATA-SET CREATION

Transaction	Seat Cover	Audio System	Car Cover	Steering Cover	Toolbox	Foot mats	Mud flaps	Window Tints
1	Y	N	Y	N	N	Y	Y	N
2	Y	Y	N	Y	N	Y	N	N
3	N	N	Y	N	N	N	N	Y
4	N	Y	N	N	Y	N	Y	Y
5	N	N	N	N	N	N	N	N
6	Y	Y	N	N	Y	Y	N	N
7	N	N	N	N	N	N	N	Y
8	N	N	N	N	N	N	Y	N
9	N	Y	N	Y	Y	Y	N	N
10	N	Y	N	N	N	N	N	Y
11	N	N	Y	N	N	Y	N	N
12	N	N	N	Y	Y	N	N	N
13	Y	N	N	N	N	N	N	N
14	N	N	N	N	N	Y	N	N
15	N	N	Y	N	N	N	N	N
Total	4	4	4	3	4	6	3	4

```

RStudio

fat_exam.Rmd
23- ## Importing the data-set
24- ```{r}
25- df = read.csv('data.csv')
26- head(df)
27- ```

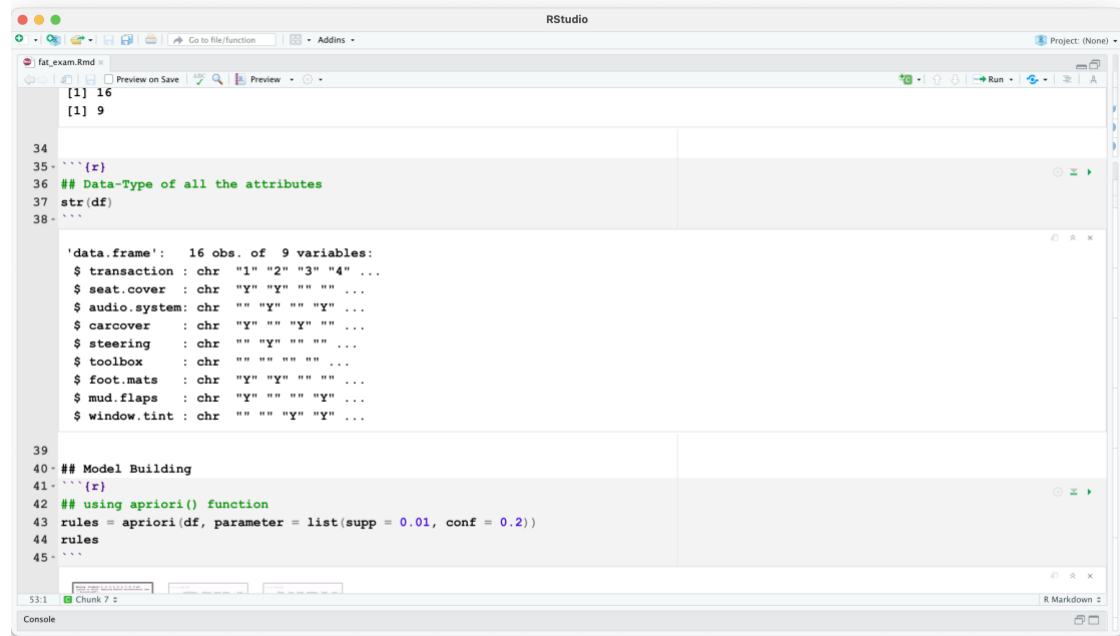
Description: df [6 x 9]
  transaction seat.cover audio.system carcover steering toolbox foot.mats mud.flaps window.tint
1 1 Y N Y N N Y Y N
2 2 Y Y N Y Y N Y N N
3 3 N N Y N N N N Y Y
4 4 N Y N N N Y N Y Y
5 5 N N N N N Y N N N
6 6 Y N N N N N Y N N
6 rows

28
29- ## Knowing about the data-set
30- ```{r}
31- nrow(df)
32- ncol(df)
33- ```

[1] 16
[1] 9

53.1 Chunk 7
Console

```



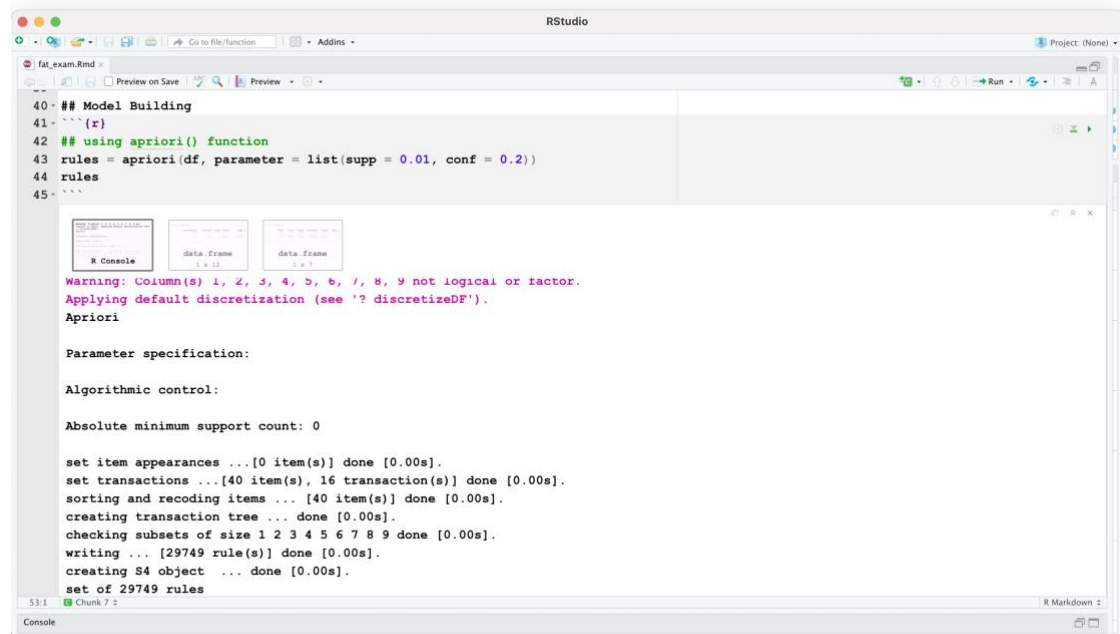
```
[1] 16
[1] 9

34
35- ``{r}
36 ## Data-Type of all the attributes
37 str(df)
38- ``

'data.frame': 16 obs. of 9 variables:
 $ transaction : chr "1" "2" "3" "4" ...
 $ seat.cover : chr "y" "y" "" "" ...
 $ audio.system: chr "" "y" "" "y" ...
 $ carcover : chr "y" "" "y" "" ...
 $ steering : chr "" "y" "" "" ...
 $ toolbox : chr "" "" "" "" ...
 $ foot.mats : chr "y" "y" "" "" ...
 $ mud.flaps : chr "y" "" "" "y" ...
 $ window.tint: chr "" "" "y" "y" ...

39
40- ## Model Building
41- ``{r}
42 ## using apriori() function
43 rules = apriori(df, parameter = list(supp = 0.01, conf = 0.2))
44 rules
45- ``
```

Model Building and Output



```
40- ## Model Building
41- ``{r}
42 ## using apriori() function
43 rules = apriori(df, parameter = list(supp = 0.01, conf = 0.2))
44 rules
45- ``

Warning: Column(s) 1, 2, 3, 4, 5, 6, 7, 8, 9 not logical or factor.
Applying default discretization (see '? discretizeDF').
Apriori

Parameter specification:

Algorithmic control:

Absolute minimum support count: 0

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[40 item(s), 16 transaction(s)] done [0.00s].
sorting and recoding items ... [40 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 7 8 9 done [0.00s].
writing ... [29749 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
set of 29749 rules
```