

L21-6205

Syed Sheraz Hussain

Lab - 4

a) What is the size of the training set?

The size of the training set is 14*5. In which 14 instances and 5 attributes.

b) How many attributes exist in the training set?

There are 5 attributes.

c) How many instances are positive (Enjoy = yes) and how many negative?

Positives instances are **9**.

Negative instances are **5**.

d) Which attribute best separates the data?

Outlook attribute best separates the data

e) How many elements from the data set have the humidity attribute set as high?

9 elements from the data set have the humidity attribute set as high

The J48 classifier is used by the Choose button:

The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected. In the 'Classifier' section, the 'Choose' button is highlighted, and the selected classifier is 'J48 -C 0.25 -M 2'. The 'Test options' section on the left has 'Use training set' selected. The 'Classifier output' section on the right displays the following results:

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	14	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	yes
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	no
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

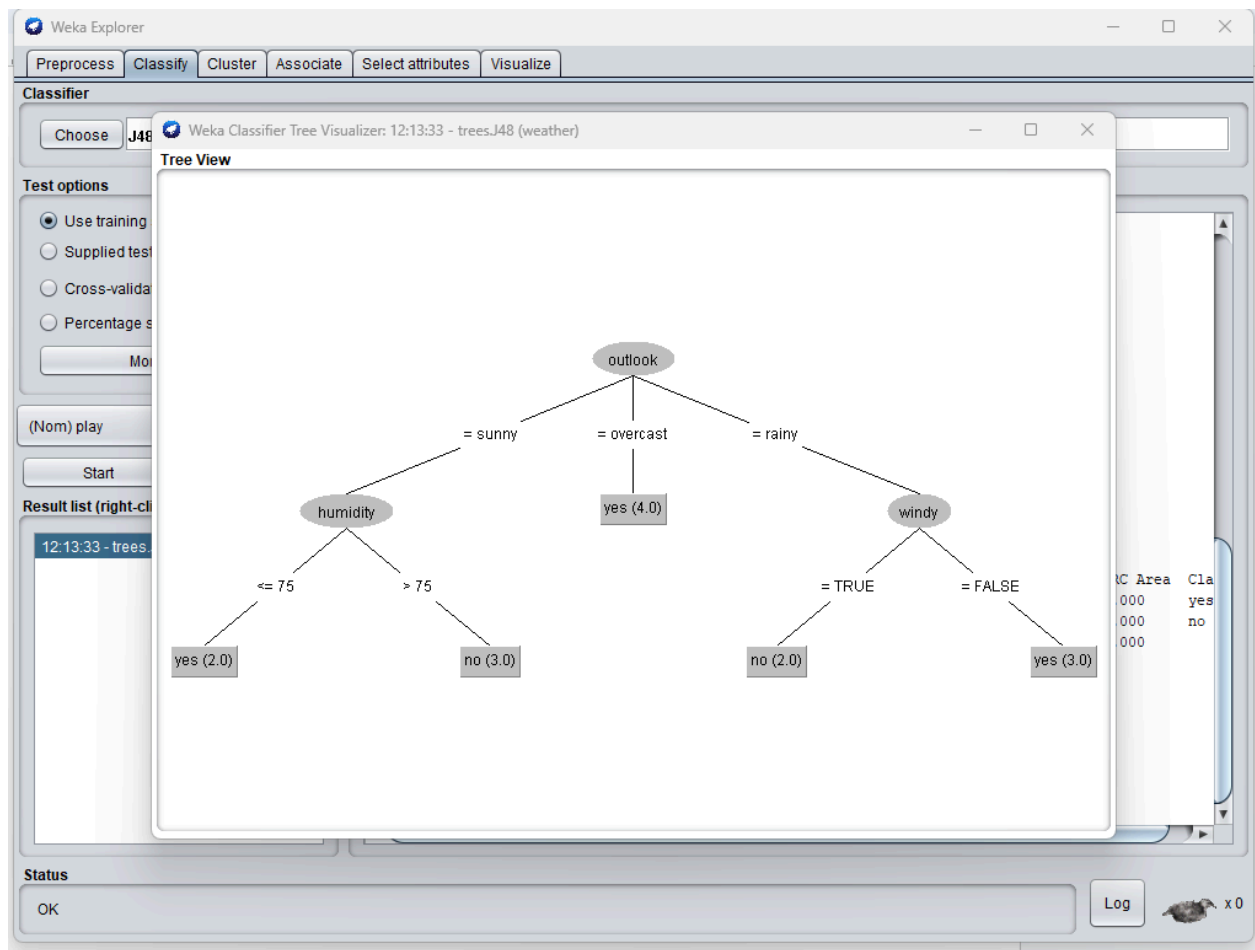
a b <-- classified as

9 0 | a = yes

0 5 | b = no

The 'Result list' on the left shows a single result: '12:13:33 - trees.J48'. The 'Status' bar at the bottom indicates 'OK'.

TREE:



Part 3:

Load the data set credit-g

And then

The J48 classifier is used by the Choose button:

The screenshot shows the Weka Explorer window with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'J48 -C 0.25 -M 2'. The 'Test options' section has 'Use training set' selected. The 'Classifier output' pane displays the following results:

```
=== Evaluation on training set ===
Time taken to test model on training data: 0.01 seconds

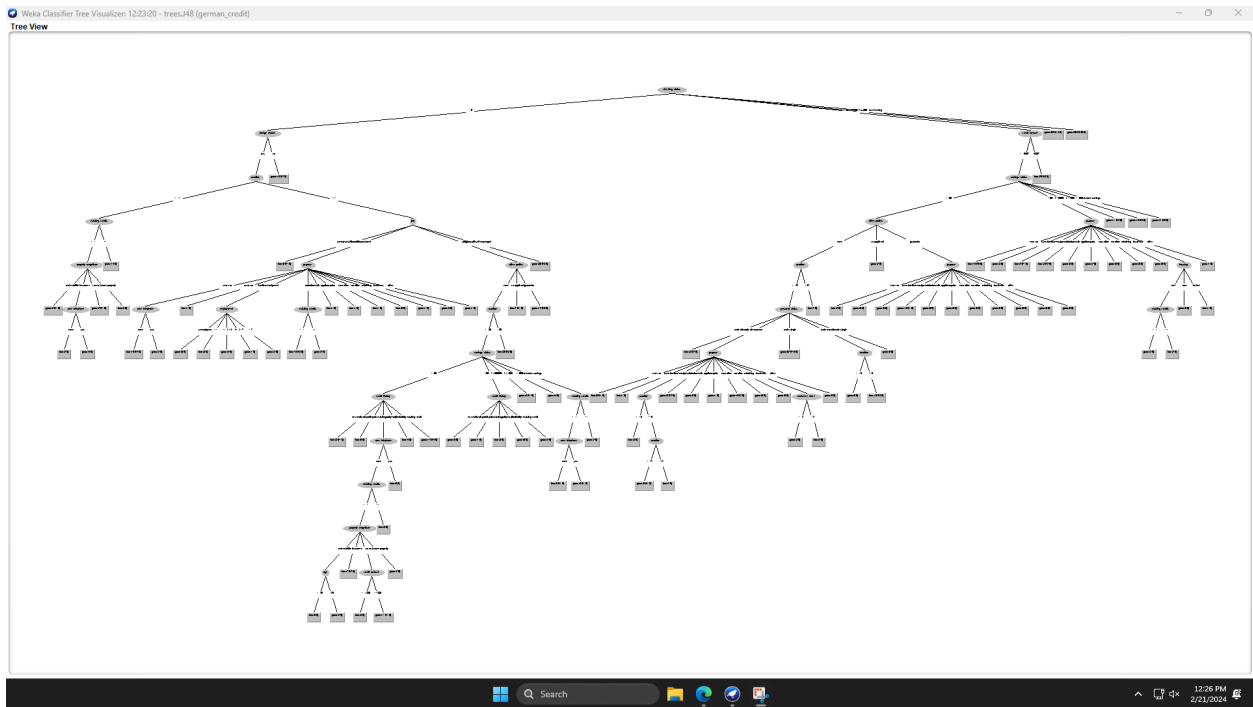
=== Summary ===
Correctly Classified Instances      855      85.5 %
Incorrectly Classified Instances    145      14.5 %
Kappa statistic                    0.6251
Mean absolute error                 0.2312
Root mean squared error             0.34
Relative absolute error             55.0377 %
Root relative squared error         74.2015 %
Total Number of Instances          1000

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
0.956    0.380    0.854    0.956    0.902    0.640    0.857    0.905    goo
0.620    0.044    0.857    0.620    0.720    0.640    0.857    0.783    bad
Weighted Avg.    0.855    0.279    0.855    0.855    0.847    0.640    0.857    0.869

=== Confusion Matrix ===
      a   b   <-- classified as
669  31 |   a = good
114 186 |   b = bad
```

The 'Result list' on the left shows two entries: '12:13:33 - trees.J48' and '12:23:20 - trees.J48', with the latter selected. The 'Status' bar at the bottom shows 'OK' and a 'Log' button.

TREE: Credit-g



Credit data (J48 classifier):

Correctly Classified Instances	855	85.5	%
Incorrectly Classified Instances	145	14.5	%

Credit data (Zero-R):

Correctly Classified Instances	700	70	%
Incorrectly Classified Instances	300	30	%

Comparing:

J48 classifier classifies better than the zero-R classifier.

Credit data (Decision Table):

Correctly Classified Instances	763	76.3	%
Incorrectly Classified Instances	237	23.7	%

Credit data (Random Forest):

Correctly Classified Instances	1000	100	%
Incorrectly Classified Instances	0	0	%

6.

J48 classifier :	Correctly Classified Instances	855	85.5	%
Zero-R :	Correctly Classified Instances	700	70	%
Random Forest:	Correctly Classified Instances	1000	100	%
Decision Table:	Correctly Classified Instances	763	76.3	%

7.

Recall of good class is higher than bad and correctly classified class is good, because if see the distribution we can see that good instances of class are 700 and of bad is 300 .and correctly classified as good are 855 and that of bad is 145 so it shows the distribution of the attributes.

8.

Data cleaning, handling missing values, and feature scaling are essential preprocessing steps. Understanding the data helps you decide how to handle outliers, impute missing values, and normalize features appropriately.