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**.

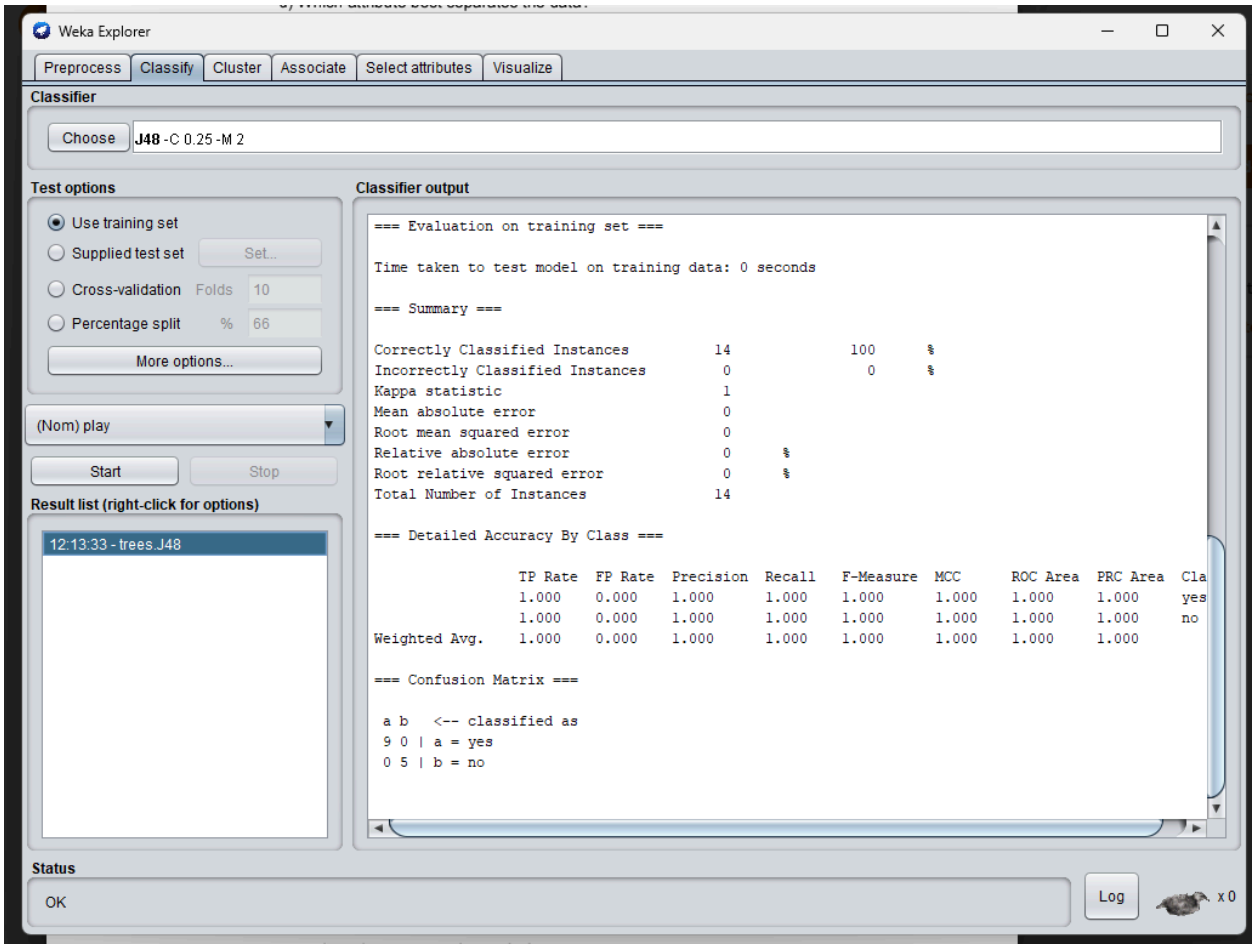
d) Which attribute best separates the data?

Outlook attribute best separates the data

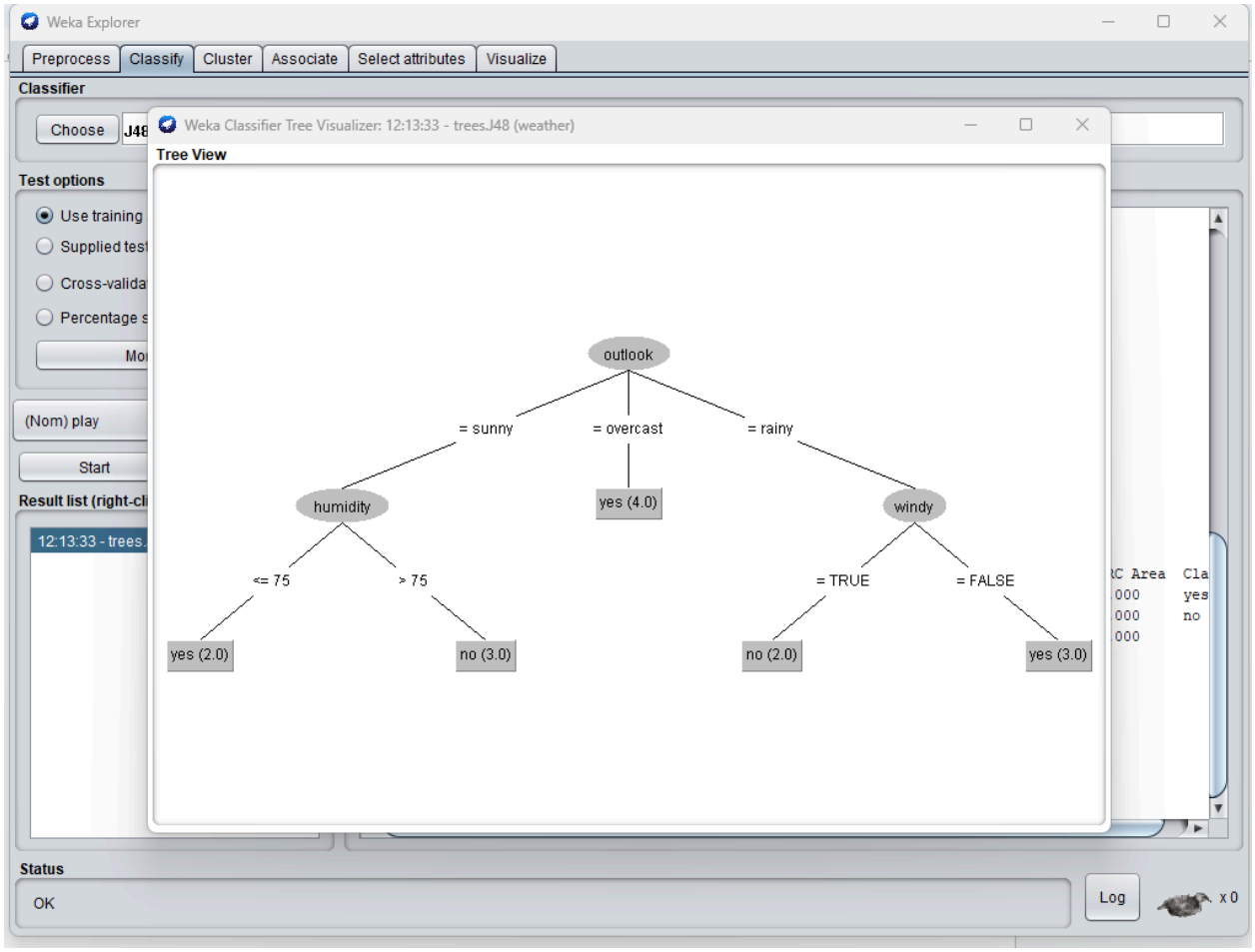
e) How many elements from the data set have the humidity attribute set as high?

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

The J48 classifier is used by the Choose button:

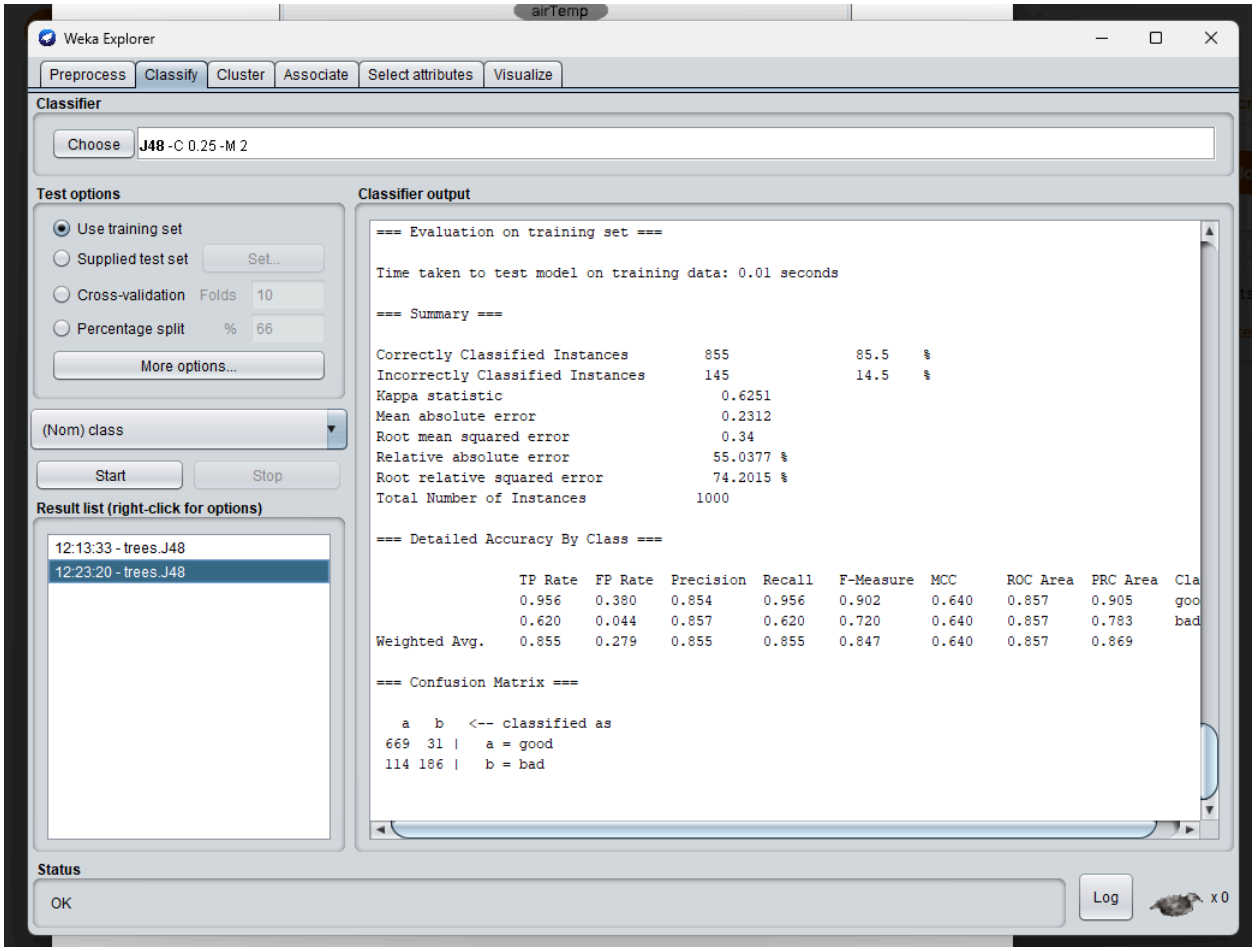


TREE:

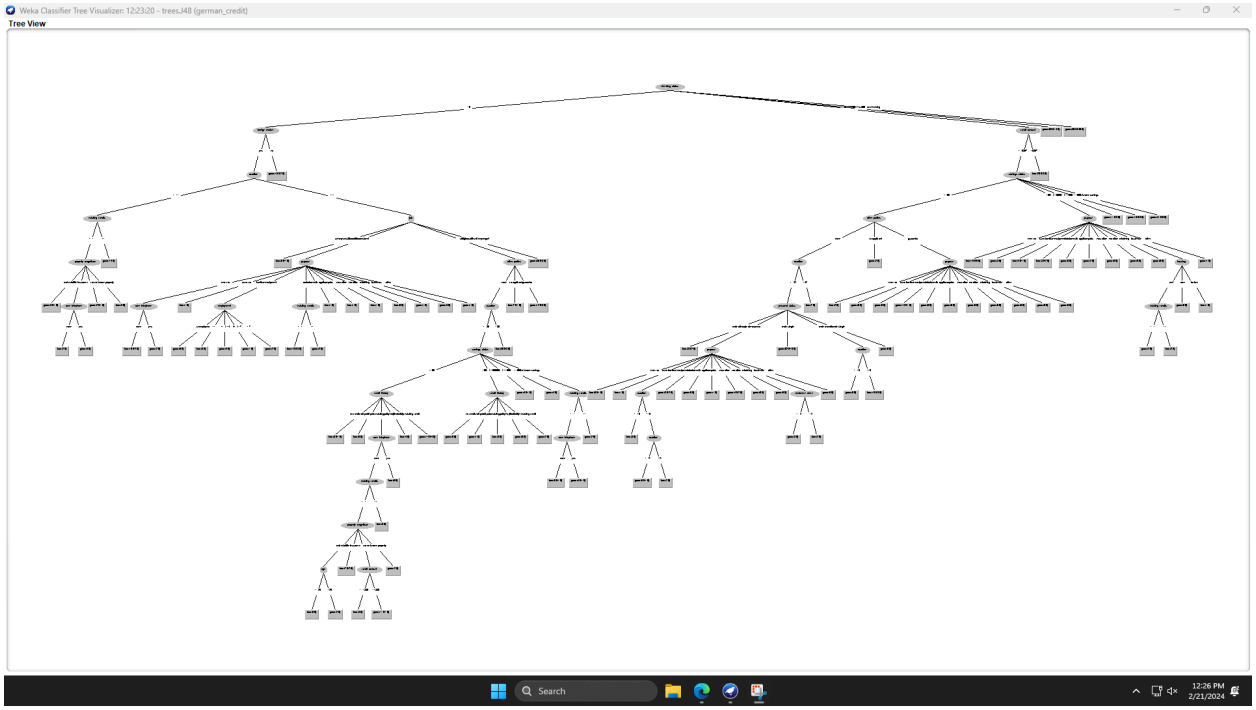

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Test options
Use training
Supplied test
Cross-valida
Percentage s
(Nom) play
Start
Result list (right-cli
12:13:33 - trees
Status
OK
Log
x 0
Weka Classifier Tree Visualizer: 12:13:33 - trees.J48 (weather)
Tree View
outlook
= sunny
= overcast
= rainy
humidity
yes (4.0)
windy
<= 75
> 75
= TRUE
= FALSE
yes (2.0)
no (3.0)
no (2.0)
yes (3.0)

Part 3:
Load the data set credit-g

And then
The J48 classifier is used by the Choose 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.