

Name: Minahil Mobin

Roll No: 21L-XXXX

Section: BDS-XXX

Data Mining Lab # 04:

Question # 01:

Part # 01 & #02:

a) What is the size of the training set?

Answer: 286

b) How many attributes exist in the training set?

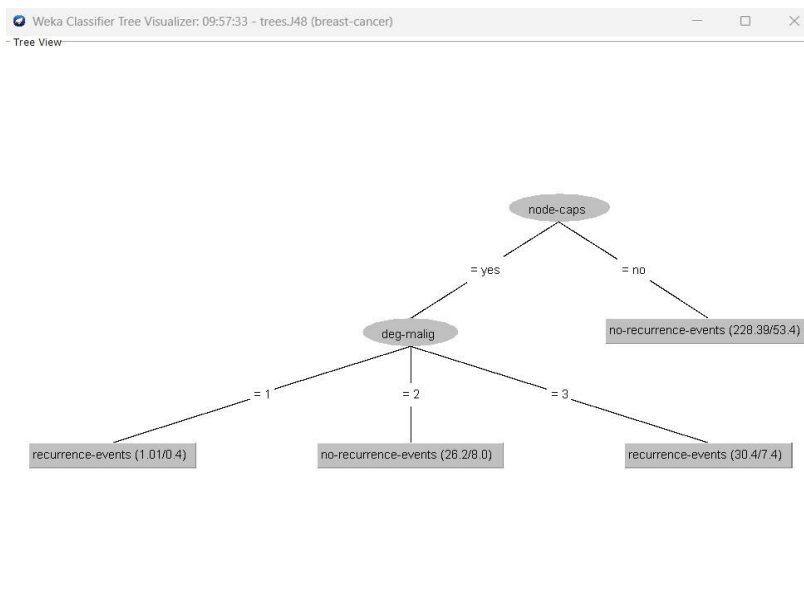
Answer: 10

c) How many instances are positive (node-caps = yes) and how many negative?

Answer: The number of positive instances is 56 and negative is 222.

d) Which attribute best separates the data?

Answer: node-caps is the attribute that best separates the data.



e) How many elements from the data set have the Recurrence as positive?

Answer: 201

Part # 03:

Question:

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' section on the left has 'Percentage split' set to 66%. The 'Classifier output' section on the right displays the following performance metrics:

Metric	Value	Percentage
Correctly Classified Instances	247	72.6471 %
Incorrectly Classified Instances	93	27.3529 %
Kappa statistic	0.2687	
Mean absolute error	0.3351	
Root mean squared error	0.4836	
Relative absolute error	80.8066 %	
Root relative squared error	108.7935 %	
Total Number of Instances	340	

Below the metrics, a 'Detailed Accuracy By Class' table is shown:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0.836	0.578	0.801	0.836	0.818	0.270
	0.422	0.164	0.481	0.422	0.450	0.270
Weighted Avg.	0.726	0.468	0.716	0.726	0.721	0.270

A 'Confusion Matrix' is also displayed:

```
a  b  <-- classified as
209 41 | a = good
52 38 | b = bad
```

The screenshot shows the Weka Explorer interface with the ZeroR classifier selected. The 'Test options' section on the left has 'Percentage split' set to 66%. The 'Classifier output' section on the right displays the following performance metrics:

Metric	Value	Percentage
Correctly Classified Instances	250	73.5294 %
Incorrectly Classified Instances	90	26.4706 %
Kappa statistic	0	
Mean absolute error	0.4147	
Root mean squared error	0.4445	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	340	

Below the metrics, a 'Detailed Accuracy By Class' table is shown:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	1.000	1.000	0.735	1.000	0.847	?
	0.000	0.000	?	0.000	?	?
Weighted Avg.	0.735	0.735	?	0.735	?	?

A 'Confusion Matrix' is also displayed:

```
a  b  <-- classified as
250 0 | a = good
90 0 | b = bad
```

The result with both J48 & ZeroR are good and pretty much the same but the result with ZeroR is slightly better than J4.

Other Classifiers:

Random Tree:

The screenshot shows the Orange3 software interface with the Random Tree classifier selected. The 'Test options' panel on the left shows 'Percentage split' at 66%. The 'Classifier output' panel on the right displays the following data:

Metric	Value	Percentage
Correctly Classified Instances	215	63.2353 %
Incorrectly Classified Instances	125	36.7647 %
Kappa statistic	0.1153	
Mean absolute error	0.3717	
Root mean squared error	0.5956	
Relative absolute error	89.634	%
Root relative squared error	133.9925	%
Total Number of Instances	340	

Below this, the 'Detailed Accuracy By Class' table is shown:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
0.712	0.589	0.771	0.712	0.740	0.116	
0.411	0.288	0.339	0.411	0.372	0.116	
Weighted Avg.	0.632	0.509	0.656	0.632	0.643	0.116

The 'Confusion Matrix' is also displayed:

a \ b	good	bad
good	178	72
bad	53	37

Random Forest:

The screenshot shows the Orange3 software interface with the Random Forest classifier selected. The 'Test options' panel on the left shows 'Percentage split' at 66%. The 'Classifier output' panel on the right displays the following data:

Metric	Value	Percentage
Correctly Classified Instances	268	78.8235 %
Incorrectly Classified Instances	72	21.1765 %
Kappa statistic	0.3812	
Mean absolute error	0.3276	
Root mean squared error	0.3898	
Relative absolute error	79.0062	%
Root relative squared error	87.6993	%
Total Number of Instances	340	

Below this, the 'Detailed Accuracy By Class' table is shown:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
0.924	0.589	0.813	0.924	0.865	0.399	
0.411	0.076	0.661	0.411	0.507	0.399	
Weighted Avg.	0.788	0.453	0.773	0.788	0.770	0.399

The 'Confusion Matrix' is also displayed:

a \ b	good	bad
good	231	19
bad	53	37

Random Forest is giving the best results compared to others. In contrast, the Random tree gives worse result compared to others.