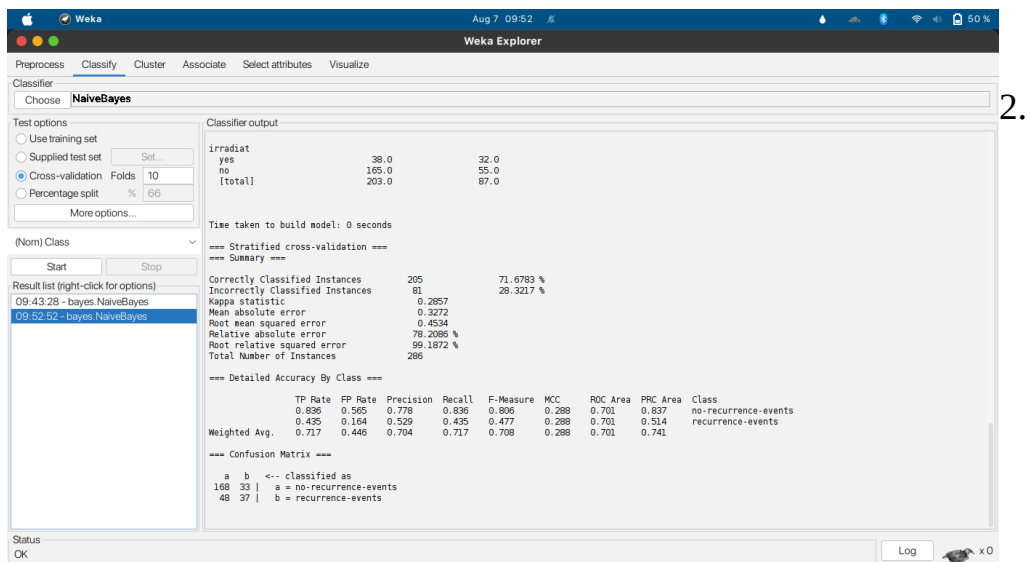


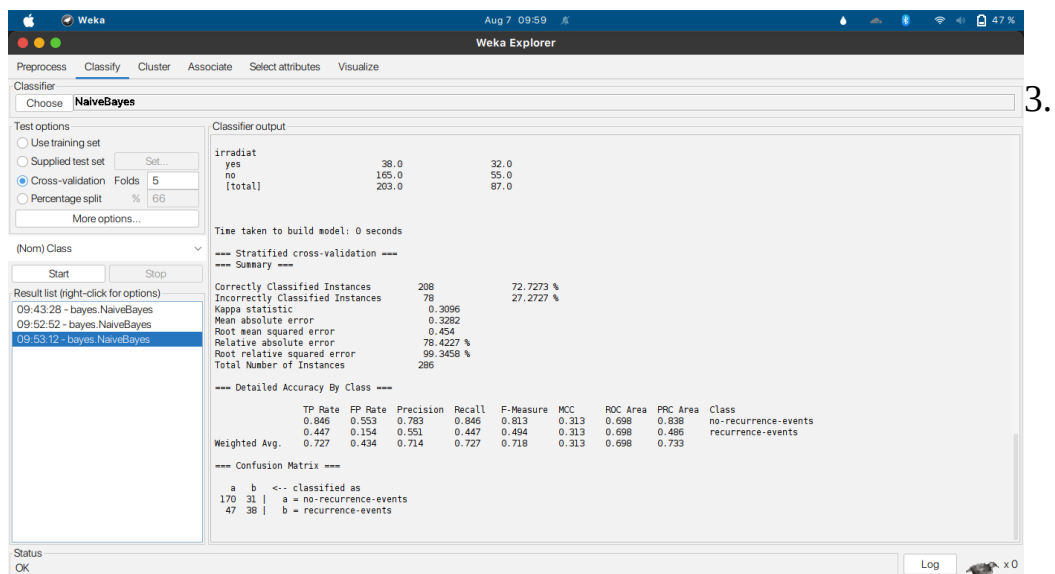
Experiment - II

Exploration of Weka tool for Breast Cancer Data classification using different Machine Learning modules

1. NaiveBayes: Cross-validation : 10 Folds



NaiveBayes: Cross-validation: 5 Folds



NaiveBayes:

Percentage Split: 80%

4.

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section has 'Percentage split' set to 80%. The 'Classifier output' section displays the following results:

no 165.0 55.0
[total] 203.0 87.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	40	70.1754 %
Incorrectly Classified Instances	17	29.8246 %
Kappa statistic	0.3257	
Mean absolute error	0.3523	
Root mean squared error	0.5058	
Relative absolute error	79.6122 %	
Root relative squared error	103.1916 %	
Total Number of Instances	57	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.702	0.392	0.692	0.702	0.691	0.332	0.619	0.679	no-recurrence-events
									recurrence-events

=== Confusion Matrix ===

a \ b	no-recurrence-events	recurrence-events
no-recurrence-events	30	6
recurrence-events	11	10

RandomTree:

Cross-validation: 10 Folds

5.

The screenshot shows the Weka Explorer interface with the RandomTree classifier selected. The 'Test options' section has 'Cross-validation' set to 10 folds. The 'Classifier output' section displays the following results:

Inv-nodes = 24-26 : recurrence-events (1/0)
Inv-nodes = 27-29 : no-recurrence-events (0/0)
Inv-nodes = 30-32 : no-recurrence-events (0/0)
Inv-nodes = 33-35 : no-recurrence-events (0/0)
Inv-nodes = 36-39 : no-recurrence-events (0/0)

Size of the tree : 444

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	191	66.7832 %
Incorrectly Classified Instances	95	33.2168 %
Kappa statistic	0.1855	
Mean absolute error	0.3533	
Root mean squared error	0.5699	
Relative absolute error	84.4448 %	
Root relative squared error	124.6637 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.668	0.487	0.660	0.668	0.664	0.186	0.588	0.629	no-recurrence-events
									recurrence-events

=== Confusion Matrix ===

a \ b	no-recurrence-events	recurrence-events
no-recurrence-events	157	44
recurrence-events	51	34

RandomTree:

Cross-validation: 5 Folds

6.

The screenshot shows the Weka Explorer interface with the RandomTree classifier selected. The test options are set to 'Percentage split' with a value of 80. The classifier output shows a size of 444 and a time taken to build the model of 0 seconds. The stratified cross-validation summary is as follows:

Summary	
Correctly Classified Instances	187
Incorrectly Classified Instances	99
Kappa statistic	0.1012
Mean absolute error	0.3548
Root mean squared error	0.574
Relative absolute error	84.7935 %
Root relative squared error	125.5996 %
Total Number of Instances	286

The detailed accuracy by class is as follows:

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.811	0.718	0.728	0.811	0.767	0.103	0.552	0.725	no-recurrence-events
0.282	0.189	0.387	0.282	0.327	0.103	0.552	0.336	recurrence-events
Weighted Avg.								
0.654	0.561	0.626	0.654	0.636	0.103	0.552	0.610	

The confusion matrix is as follows:

Confusion Matrix		
a	b	<- classified as
163	38	a = no-recurrence-events
61	24	b = recurrence-events

RandomTree:
PercentageSplit: 80%

The screenshot shows the Weka Explorer interface with the RandomTree classifier selected. The test options are set to 'Cross-validation' with a value of 10 folds. The classifier output shows a size of 444 and a time taken to build the model of 0 seconds. The evaluation on test split summary is as follows:

Summary	
Correctly Classified Instances	39
Incorrectly Classified Instances	18
Kappa statistic	0.212
Mean absolute error	0.3005
Root mean squared error	0.5123
Relative absolute error	67.9203 %
Root relative squared error	104.5244 %
Total Number of Instances	57

The detailed accuracy by class is as follows:

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.944	0.762	0.680	0.944	0.791	0.268	0.706	0.756	no-recurrence-events
0.238	0.056	0.714	0.238	0.357	0.268	0.706	0.554	recurrence-events
Weighted Avg.								
0.684	0.502	0.693	0.684	0.631	0.268	0.706	0.681	

The confusion matrix is as follows:

Confusion Matrix		
a	b	<- classified as
34	2	a = no-recurrence-events
16	5	b = recurrence-events

7. RandomForest:
Cross-validation: 10 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose RandomForest -P100 -I100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds 10
- ☐ Percentage split % 80

More options...

(Nom) Class: Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest

Classifier output:

```

=== Classifier model (full training set) ===
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.1 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      199      69.5804 %
Incorrectly Classified Instances    87      30.4196 %
Kappa statistic                    0.1736
Mean absolute error                 0.3727
Root mean squared error             0.4613
Relative absolute error             89.0857 %
Root relative squared error        100.9171 %
Total Number of Instances          286

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
Weighted Avg.   0.671   0.129   0.480     0.282   0.356     0.184   0.634   0.409   no-recurrence-events
                  0.696   0.543   0.664     0.696   0.669     0.184   0.634   0.682   recurrence-events

=== Confusion Matrix ===
  a  b  <-- classified as
175 26 | a = no-recurrence-events
 61 24 | b = recurrence-events

```

Status: OK Log x 0

8. RandomForest: Cross-validation: 5 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose RandomForest -P100 -I100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds 5
- ☐ Percentage split % 80

More options...

(Nom) Class: Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest

Classifier output:

```

=== Classifier model (full training set) ===
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      193      67.4825 %
Incorrectly Classified Instances    93      32.5175 %
Kappa statistic                    0.1494
Mean absolute error                 0.3717
Root mean squared error             0.4652
Relative absolute error             88.8197 %
Root relative squared error        101.7892 %
Total Number of Instances          286

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
Weighted Avg.   0.675   0.538   0.648     0.675   0.656     0.153   0.636   0.689   no-recurrence-events
                  0.306   0.169   0.433     0.306   0.359     0.153   0.636   0.408   recurrence-events

=== Confusion Matrix ===
  a  b  <-- classified as
167 34 | a = no-recurrence-events
 59 26 | b = recurrence-events

```

Status: OK Log x 0

9. RandomForest: PercentageSplit: 80%

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 5
- ☒ Percentage split % 80

More options...

(Nom) Class: (Nom) Class

Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest

Classifier output:

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.04 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	41	71.9298 %
Incorrectly Classified Instances	16	28.0702 %
Kappa statistic	0.3304	
Mean absolute error	0.3632	
Root mean squared error	0.4661	
Relative absolute error	82.0776 %	
Root relative squared error	95.0847 %	
Total Number of Instances	57	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.381	0.083	0.727	0.381	0.500	0.364	0.680	0.621	no-recurrence-events
	0.719	0.422	0.721	0.719	0.693	0.364	0.680	0.721	recurrence-events

=== Confusion Matrix ===

```

a b <-- classified as
33 3 | a = no-recurrence-events
13 8 | b = recurrence-events

```

Status: OK

Log x 0

10. NaiveBayesUpdateable

Cross-validation: 10 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayesUpdateable**

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds: 10
- ☐ Percentage split % 80

More options...

(Nom) Class: (Nom) Class

Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest
- 10:11:51 - bayes.NaiveBayesUpdateable

Classifier output:

irradiat

Class	Count	Percentage
yes	38.0	32.0
no	165.0	55.0
[total]	203.0	87.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	205	71.6783 %
Incorrectly Classified Instances	81	28.3217 %
Kappa statistic	0.2857	
Mean absolute error	0.3272	
Root mean squared error	0.4534	
Relative absolute error	78.2086 %	
Root relative squared error	99.1872 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.836	0.565	0.778	0.836	0.806	0.288	0.701	0.837	no-recurrence-events
	0.435	0.164	0.529	0.435	0.477	0.288	0.701	0.514	recurrence-events

=== Confusion Matrix ===

```

a b <-- classified as
168 33 | a = no-recurrence-events
48 37 | b = recurrence-events

```

Status: OK

Log x 0

11. NaiveBayes Updateable

Cross-validation: 5 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayesUpdateable

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds 5
- ☐ Percentage split % 80

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest
- 10:11:51 - bayes.NaiveBayesUpdateable
- 10:13:32 - bayes.NaiveBayesUpdateable

Classifier output

```

irradiat
yes      38.0      32.0
no       165.0     55.0
[total]  203.0     87.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      208      72.7273 %
Incorrectly Classified Instances    78      27.2727 %
Kappa statistic                    0.3096
Mean absolute error                 0.3282
Root mean squared error             0.454
Relative absolute error             78.4227 %
Root relative squared error        99.3458 %
Total Number of Instances          286

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.447   0.154   0.551     0.447   0.494     0.313   0.698   0.486   no-recurrence-events
                0.727   0.434   0.714     0.727   0.718     0.313   0.698   0.733   recurrence-events

=== Confusion Matrix ===
  a  b  <-- classified as
170 31 | a = no-recurrence-events
 47 38 | b = recurrence-events

```

Status: OK

Log

12. NaiveBayes Updateable Percentage Split: 80%

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayesUpdateable

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds 5
- ☒ Percentage split % 80

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest
- 10:11:51 - bayes.NaiveBayesUpdateable
- 10:13:32 - bayes.NaiveBayesUpdateable
- 10:14:07 - bayes.NaiveBayesUpdateable

Classifier output

```

no      165.0     55.0
[total] 203.0     87.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      40      70.1754 %
Incorrectly Classified Instances    17      29.8246 %
Kappa statistic                    0.3257
Mean absolute error                 0.3523
Root mean squared error             0.5058
Relative absolute error             79.6122 %
Root relative squared error        103.1916 %
Total Number of Instances          57

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.702   0.392   0.692     0.702   0.691     0.332   0.619   0.679   no-recurrence-events
                0.476   0.167   0.625     0.476   0.541     0.332   0.619   0.584   recurrence-events

=== Confusion Matrix ===
  a  b  <-- classified as
 30  6 | a = no-recurrence-events
 11 10 | b = recurrence-events

```

Status: OK

Log

13. REP Tree: Cross-validation: 10 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds 10
- ☐ Percentage split % 80

(Nom) Class: Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest
- 10:11:51 - bayes.NaiveBayesUpdateable
- 10:13:32 - bayes.NaiveBayesUpdateable
- 10:14:07 - bayes.NaiveBayesUpdateable
- 10:14:52 - trees.REPTree

Classifier output:

```

| age = 90-99 : no-recurrence-events (0/0) [0/0]
tumor-size = 40-44 : no-recurrence-events (17/4) [5/2]
tumor-size = 45-49 : no-recurrence-events (2/1) [1/0]
tumor-size = 50-54 : no-recurrence-events (4/0) [4/3]
tumor-size = 55-59 : no-recurrence-events (0/0) [0/0]

Size of the tree : 32
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      202      70.6294 %
Incorrectly Classified Instances    84      29.3706 %
Kappa statistic                    0.1601
Mean absolute error                 0.3797
Root mean squared error             0.4652
Relative absolute error             90.7409 %
Root relative squared error        101.7845 %
Total Number of Instances          286

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
Weighted Avg.   0.706   0.572   0.669     0.706   0.664     0.182   0.621   0.661
no-recurrence-events
recurrence-events

=== Confusion Matrix ===
  a b  <-- classified as
183 18 | a = no-recurrence-events
 66 19 | b = recurrence-events

```

Status: OK Log x 0

14. REP Tree: Cross-validation: 5 Folds

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Test options:

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds 5
- ☐ Percentage split % 80

(Nom) Class: Start Stop

Result list (right-click for options):

- 09:43:28 - bayes.NaiveBayes
- 09:52:52 - bayes.NaiveBayes
- 09:53:12 - bayes.NaiveBayes
- 10:01:06 - bayes.NaiveBayes
- 10:04:27 - trees.RandomTree
- 10:06:19 - trees.RandomTree
- 10:07:27 - trees.RandomTree
- 10:07:35 - trees.RandomTree
- 10:09:01 - trees.RandomForest
- 10:09:31 - trees.RandomForest
- 10:10:42 - trees.RandomForest
- 10:11:51 - bayes.NaiveBayesUpdateable
- 10:13:32 - bayes.NaiveBayesUpdateable
- 10:14:07 - bayes.NaiveBayesUpdateable
- 10:14:52 - trees.REPTree
- 10:15:32 - trees.REPTree

Classifier output:

```

| age = 90-99 : no-recurrence-events (0/0) [0/0]
tumor-size = 40-44 : no-recurrence-events (17/4) [5/2]
tumor-size = 45-49 : no-recurrence-events (2/1) [1/0]
tumor-size = 50-54 : no-recurrence-events (4/0) [4/3]
tumor-size = 55-59 : no-recurrence-events (0/0) [0/0]

Size of the tree : 32
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      193      67.4825 %
Incorrectly Classified Instances    93      32.5175 %
Kappa statistic                    0.0738
Mean absolute error                 0.4016
Root mean squared error             0.4724
Relative absolute error             95.9705 %
Root relative squared error        103.3728 %
Total Number of Instances          286

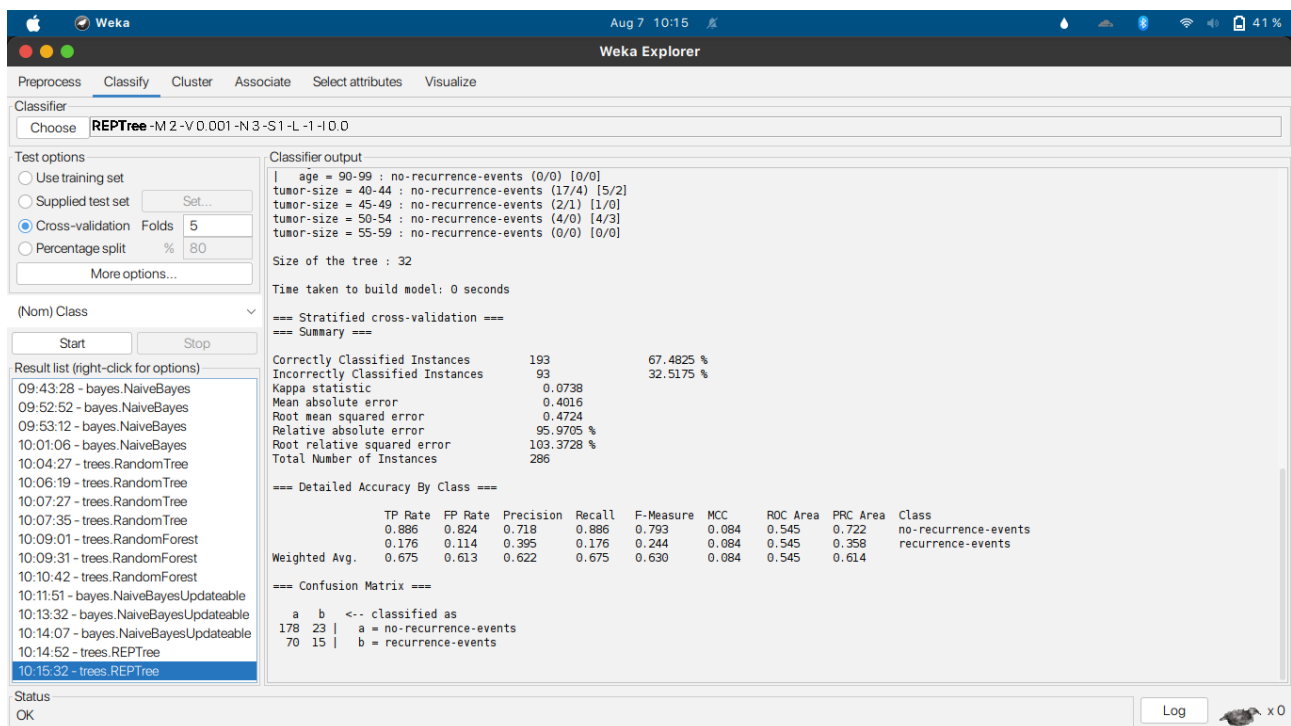
=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
Weighted Avg.   0.675   0.613   0.622     0.675   0.630     0.084   0.545   0.614
no-recurrence-events
recurrence-events

=== Confusion Matrix ===
  a b  <-- classified as
178 23 | a = no-recurrence-events
 70 15 | b = recurrence-events

```

Status: OK Log x 0

15. REP Tree: Percentage Split: 80%



DEFINITIONS:

TP Rate: rate of true positives (instances correctly classified as a given class)

FP Rate: rate of false positives (instances falsely classified as a given class)

Precision: proportion of instances that are truly of a class divided by the total instances classified as that class

Recall: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate)

F-Measure: A combined measure for precision and recall calculated as $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$

MCC is used in machine learning as a measure of the quality of binary (two-class) classifications. It takes into account true and false positives and

negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes

ROC(Receiver Operating Characteristics) area measurement: One of the most important values output by Weka. They give you an idea of how the classifiers are performing in general