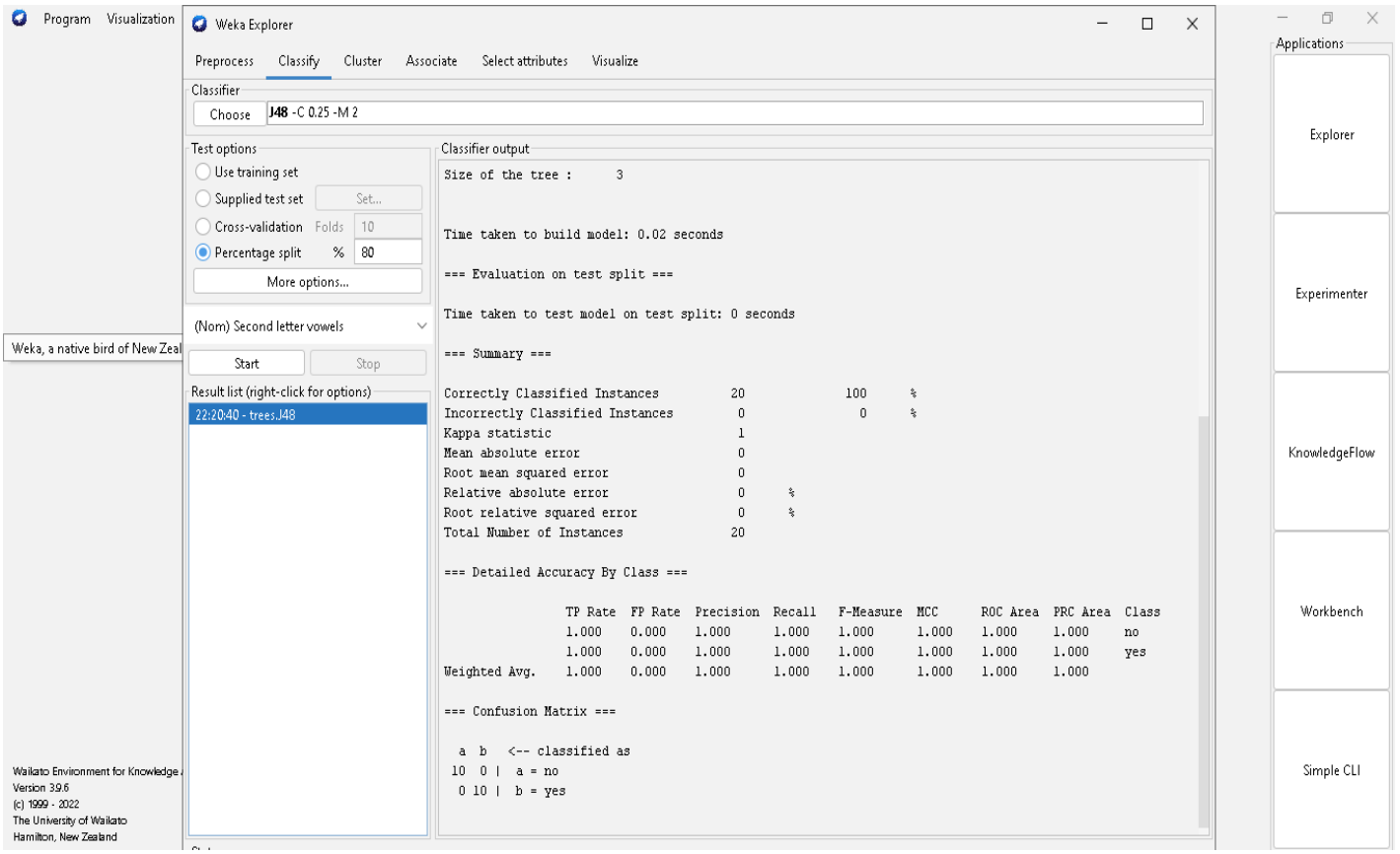
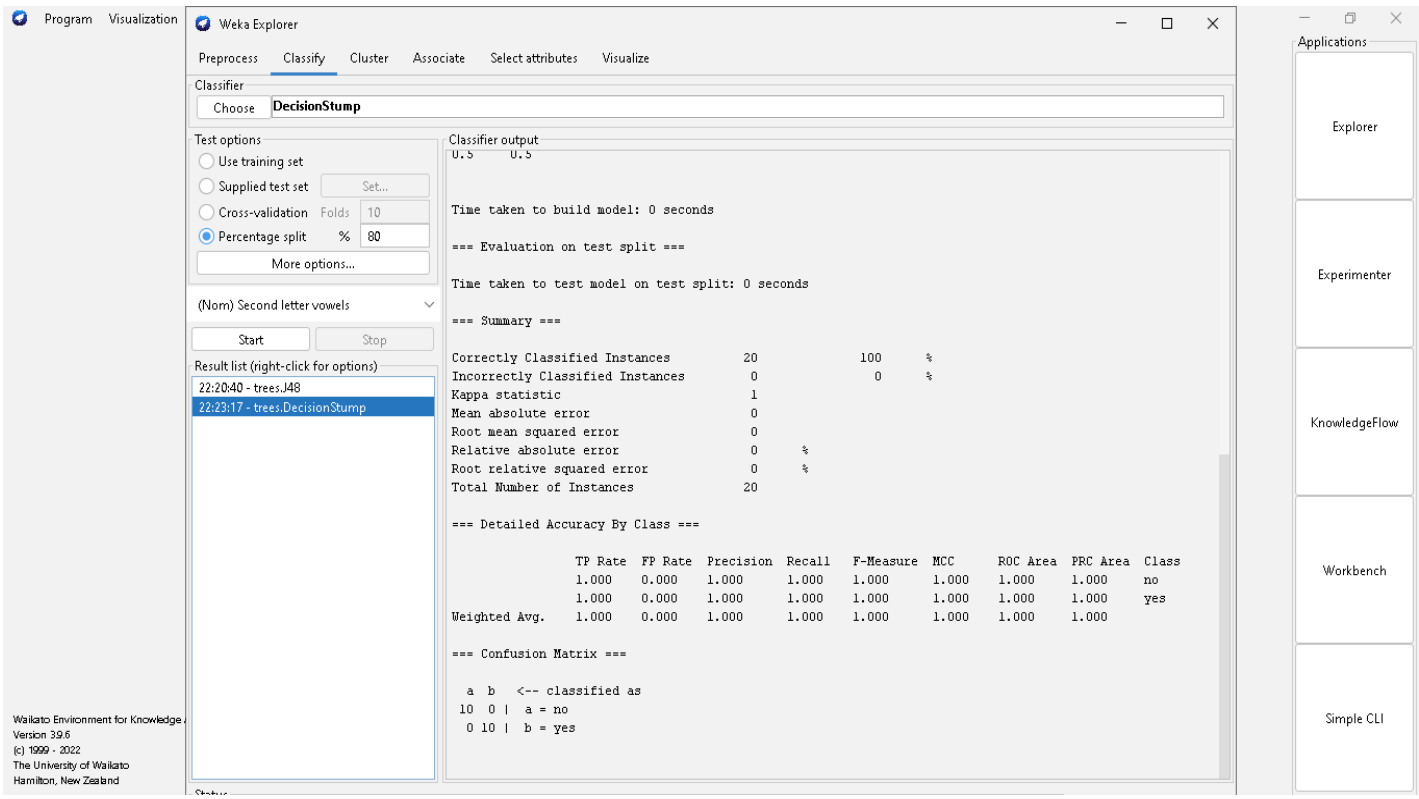
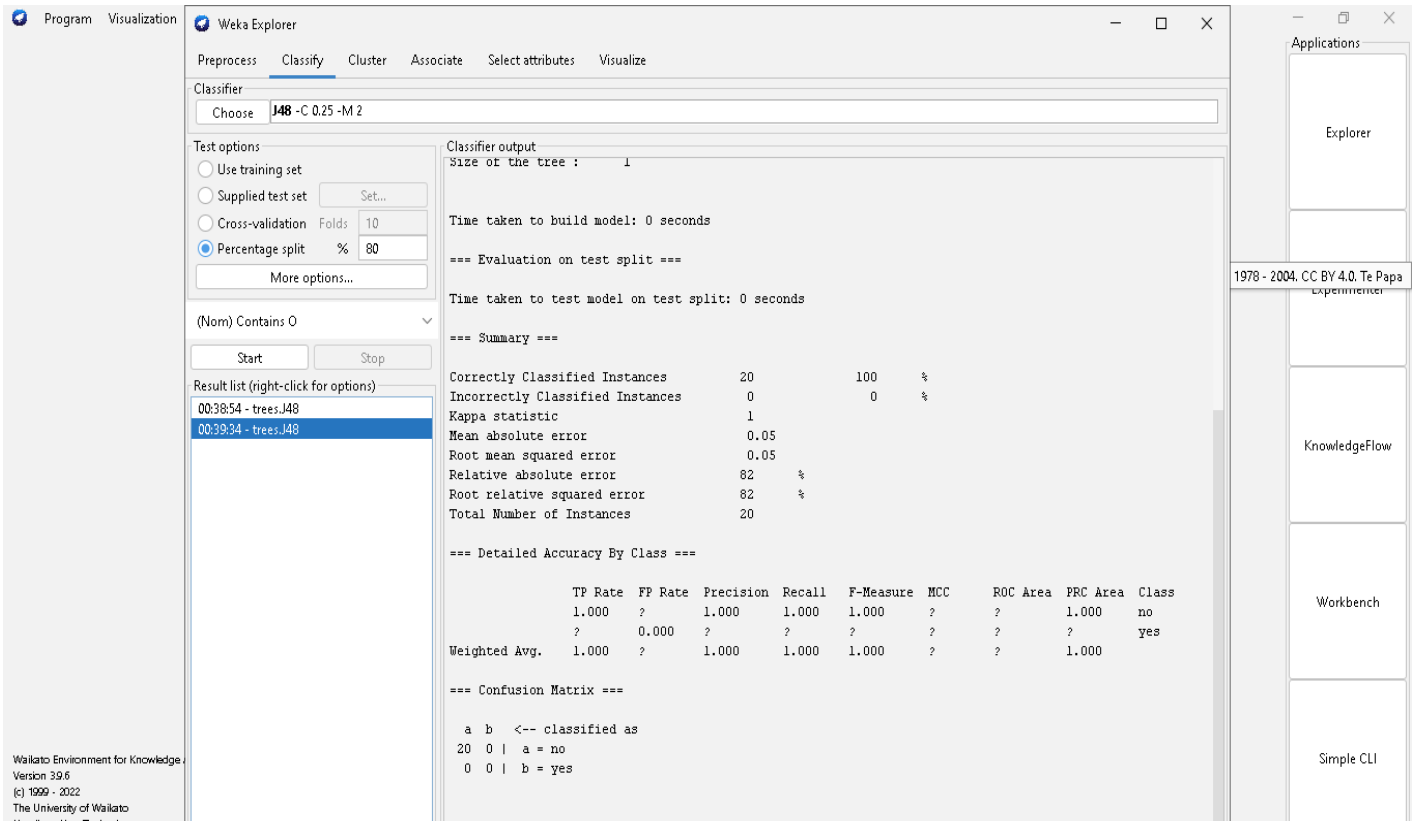
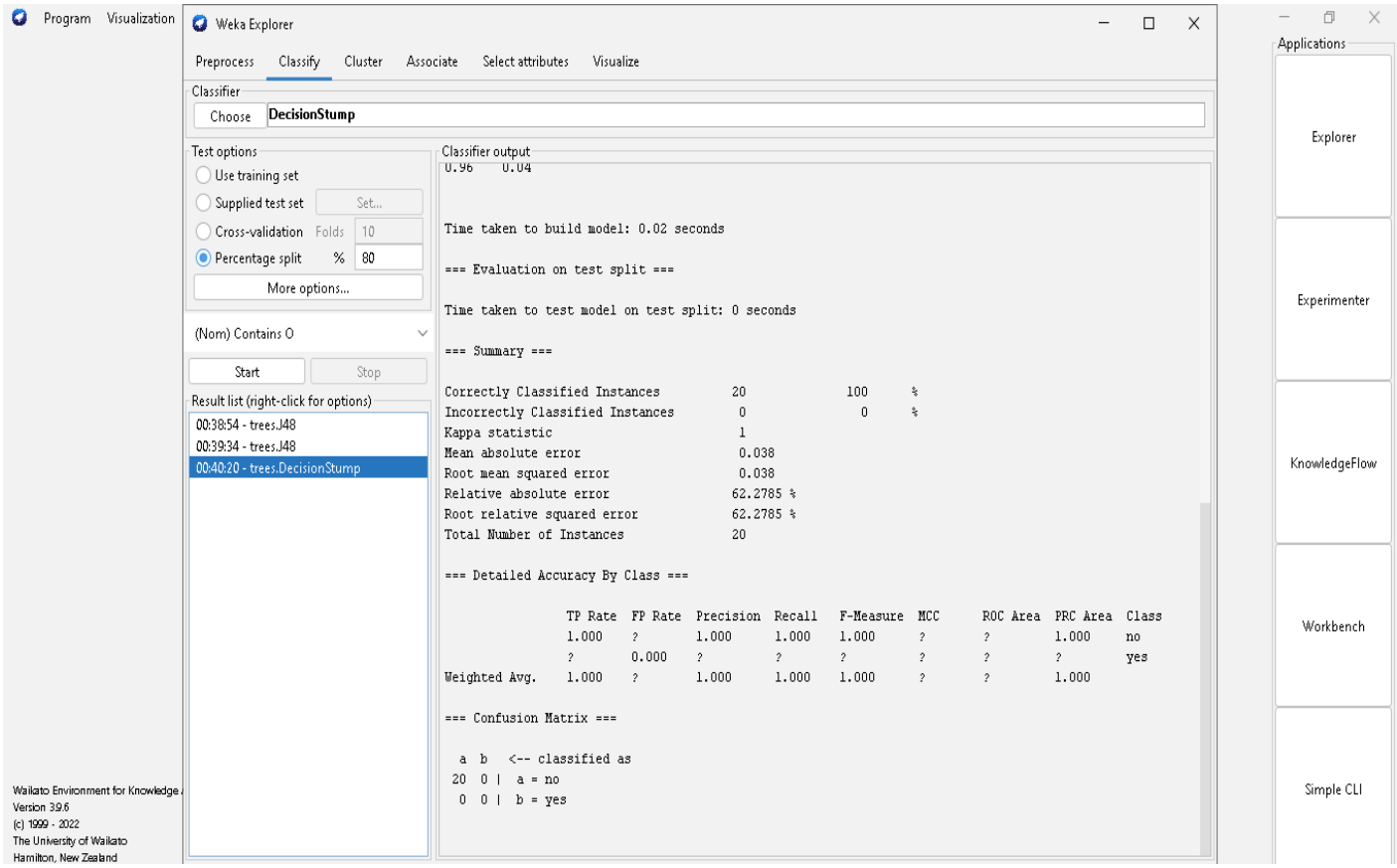

Program Visualization
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 80
More options...
(Nom) Second letter vowels
Start Stop
Result list (right-click for options)
22:20:40 - trees.J48
Classifier output
Size of the tree : 3
Time taken to build model: 0.02 seconds
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds
=== Summary ===
Correctly Classified Instances 20 100 %
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 20
=== 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 no
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 yes
Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000
=== Confusion Matrix ===
a b <-- classified as
10 0 | a = no
0 10 | b = yes
Weka, a native bird of New Zeal
Waikato Environment for Knowledge
Version 3.9.6
(c) 1999 - 2022
The University of Waikato
Hamilton, New Zealand
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI


Program Visualization
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose DecisionStump
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 80
More options...
(Nom) Second letter vowels
Start Stop
Result list (right-click for options)
22:20:40 - trees.J48
22:23:17 - trees.DecisionStump
Classifier output
0.5 0.5
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 20 100 %
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 20
=== 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 no
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 yes
Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000
=== Confusion Matrix ===
a b <-- classified as
10 0 | a = no
0 10 | b = yes
Waikato Environment for Knowledge
Version 3.9.6
(c) 1999 - 2022
The University of Waikato
Hamilton, New Zealand
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI


Program Visualization
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 80
More options...
(Nom) Contains O
Start Stop
Result list (right-click for options)
00:38:54 - trees.J48
00:39:34 - trees.J48
Classifier output
Size of the tree : 1
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 20 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0.05
Root mean squared error 0.05
Relative absolute error 82 %
Root relative squared error 82 %
Total Number of Instances 20
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
1.000 ? 1.000 1.000 1.000 ? ? 1.000 no
? 0.000 ? ? ? ? ? ? yes
Weighted Avg. 1.000 ? 1.000 1.000 1.000 ? ? 1.000
=== Confusion Matrix ===
a b <-- classified as
20 0 | a = no
0 0 | b = yes
Waikato Environment for Knowledge
Version 3.9.6
(c) 1999 - 2022
The University of Waikato
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI
1978 - 2004, CC BY 4.0, Te Papa


Program Visualization
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose DecisionStump
Test options
Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 80
More options...
(Nom) Contains O
Start Stop
Result list (right-click for options)
00:38:54 - trees.J48
00:39:34 - trees.J48
00:40:20 - trees.DecisionStump
Classifier output
0.96 0.04
Time taken to build model: 0.02 seconds
=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds
=== Summary ===
Correctly Classified Instances 20 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0.038
Root mean squared error 0.038
Relative absolute error 62.2785 %
Root relative squared error 62.2785 %
Total Number of Instances 20
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
1.000 ? 1.000 1.000 1.000 ? ? 1.000 no
? 0.000 ? ? ? ? ? ? yes
Weighted Avg. 1.000 ? 1.000 1.000 1.000 ? ? 1.000
=== Confusion Matrix ===
a b <-- classified as
20 0 | a = no
0 0 | b = yes
Waikato Environment for Knowledge
Version 3.9.6
(c) 1999 - 2022
The University of Waikato
Hamilton, New Zealand
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI

In my experience of working with a standard ML pipeline in WEKA, the process begins by carefully preparing the dataset, which includes manual feature engineering, such as adding the "Second letter vowels" and "Contains O" features to the original dataset. Once the data is prepped and converted to the ARFF format, I load it into WEKA's workbench to analyze the different characteristics of the data through visualizations and statistics offered by the tool. This step provides valuable insights, such as identifying patterns and trends within the dataset, and reveals interesting features that may impact the classification results.

After running the J48 algorithm, I could see how these new features impacted the classification outcomes. WEKA's ability to visualize the decision tree structure made it easier to understand how the model arrived at its conclusions. This process underscored how essential it is to experiment with different features, analyze the results, and adjust the data as needed to optimize model performance.