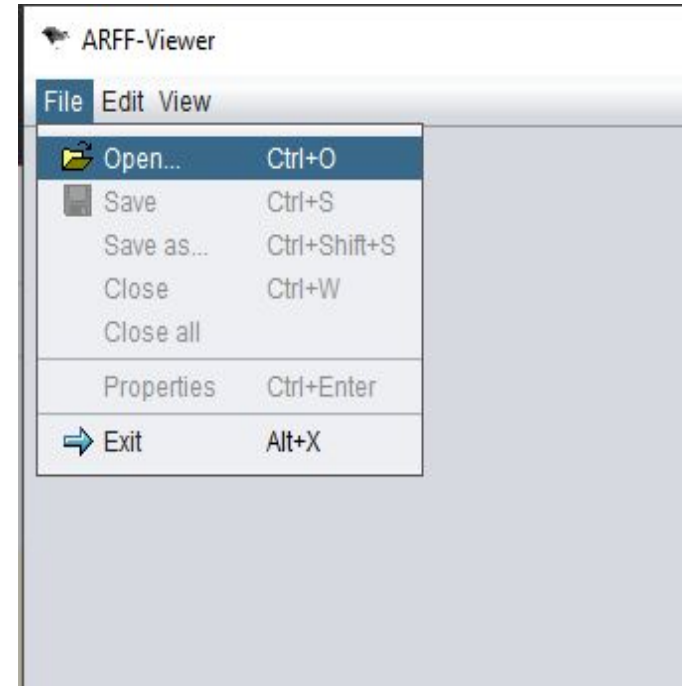
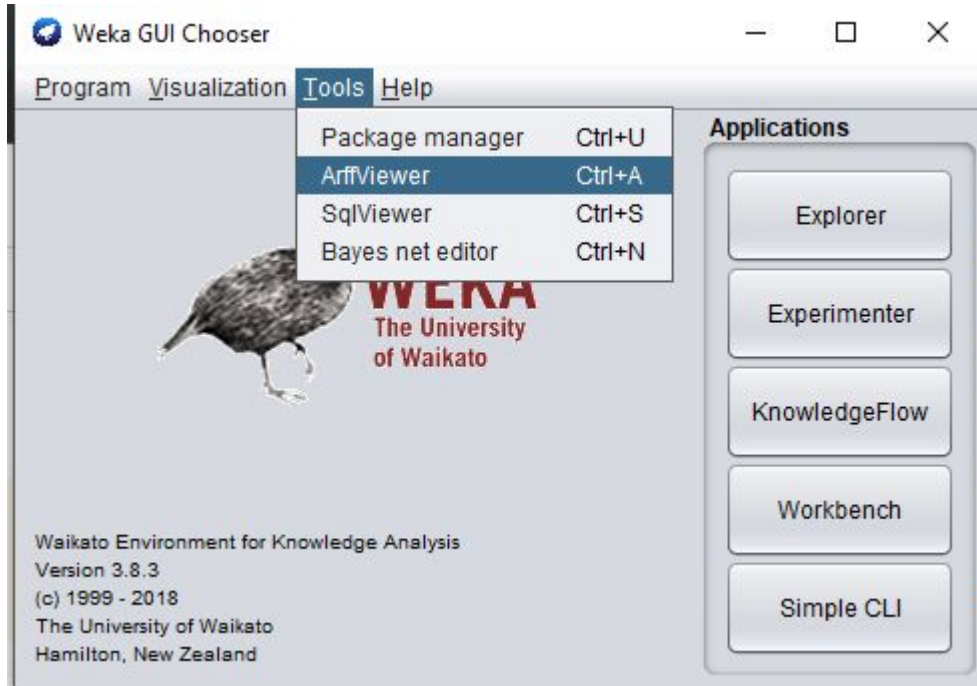


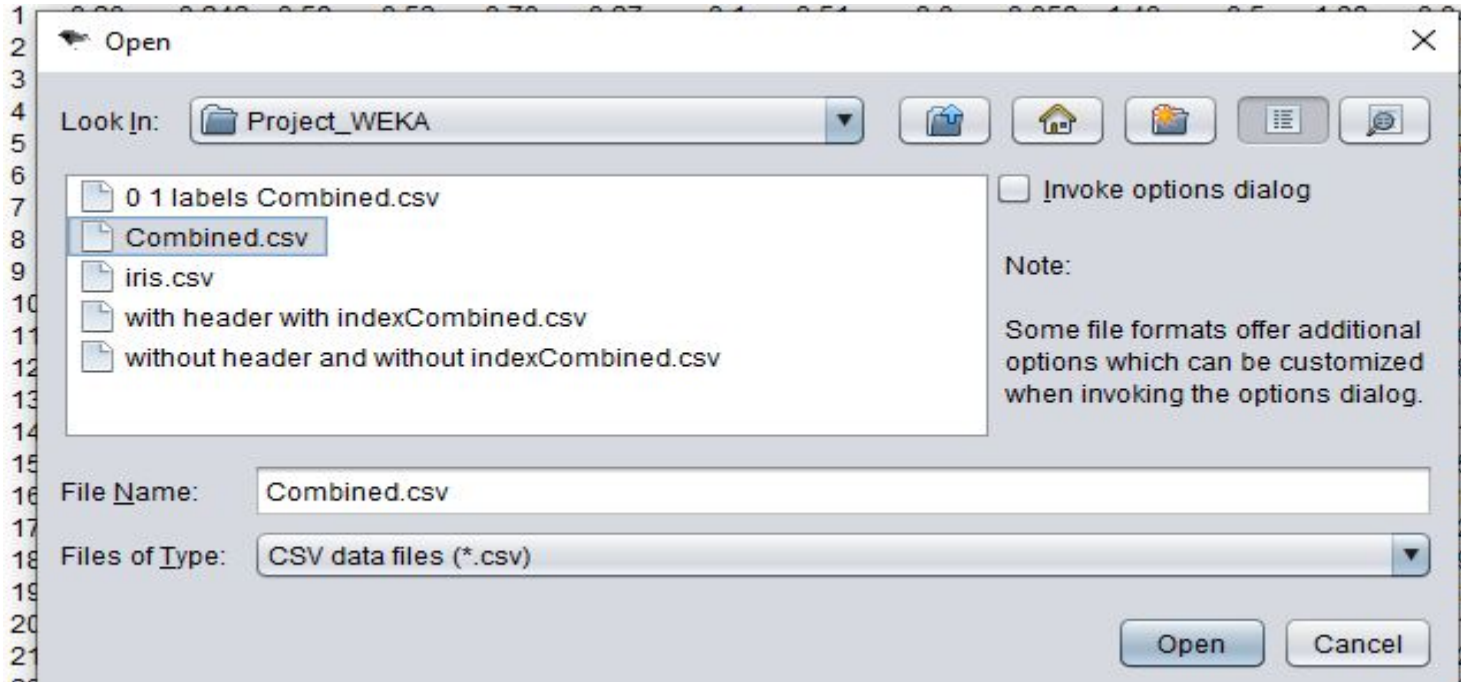
Research Paper - WEKA

Screenshots

Dataset - Compatibility check



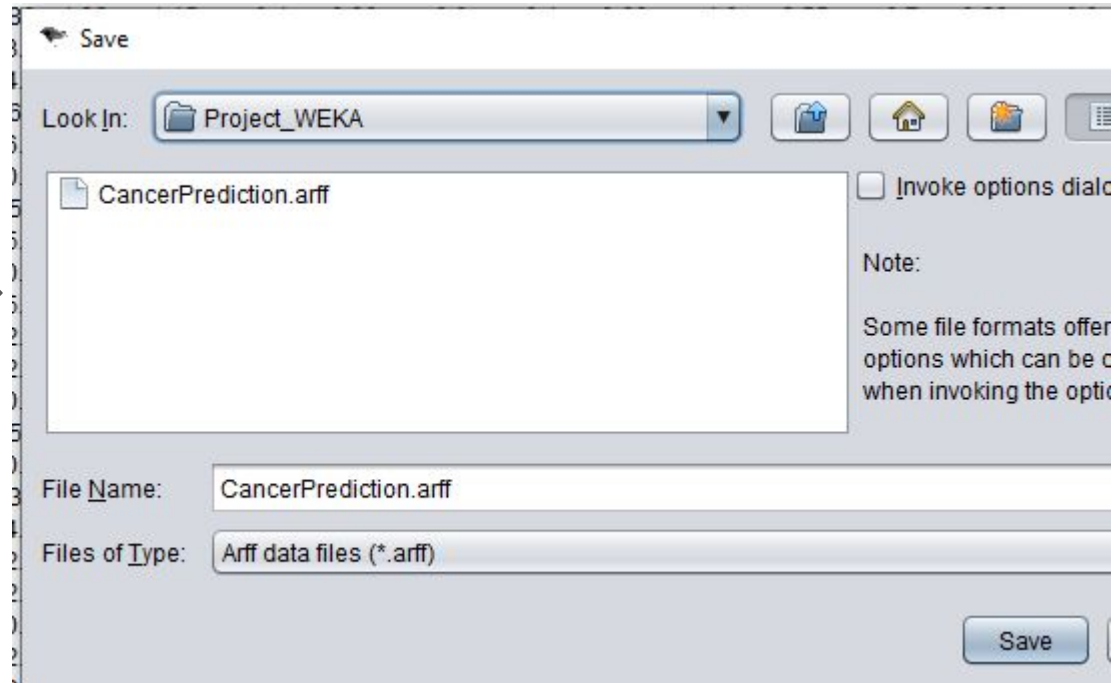
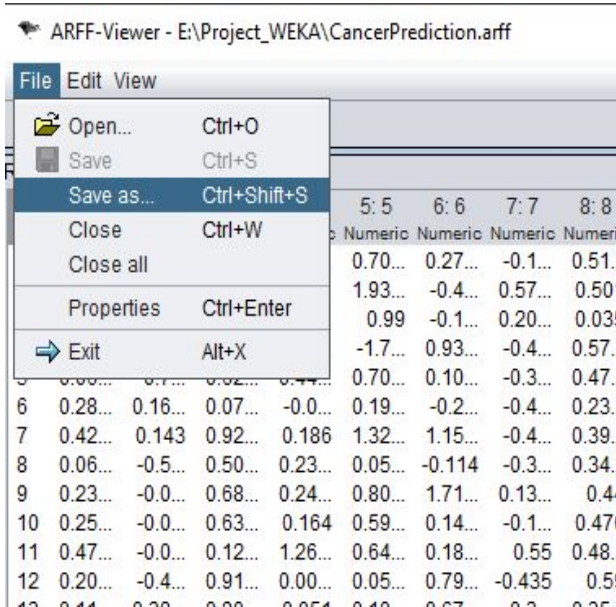
Dataset Loading.....



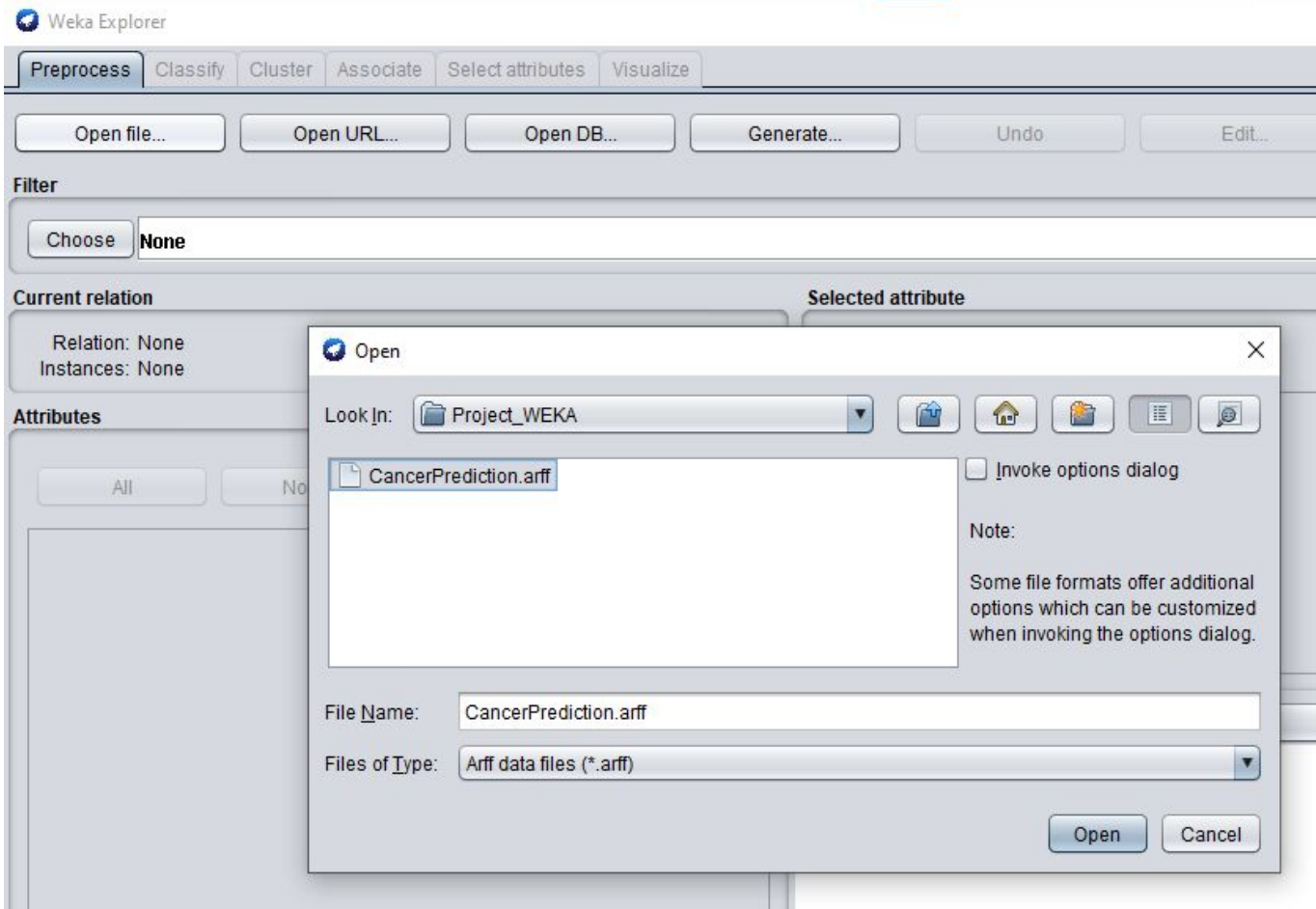
Dataset Loaded in ARFF Viewer

ARFF-Viewer - E:\Project_WEKA\CancerPrediction.arff																									
File Edit View																									
CancerPrediction.arff																									
Relation: Combined																									
No.	1: 1	2: 2	3: 3	4: 4	5: 5	6: 6	7: 7	8: 8	9: 9	10: 10	11: 11	12: 12	13: 13	14: 14	15: 15	16: 16	17: 17	18: 18	19: 19	20: 20	21: 21	22:			
	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nur		
1	0.20...	-0.242	0.59...	0.53...	0.70...	0.27...	-0.1...	0.51...	-0.0...	-0.059	1.40...	-0.5...	1.89...	0.04...	0.33...	0.27...	-2.0...	0.31...	-0.1...	-0.1...	-0.3...	0.			
2	0.86...	0.87...	-0.0...	0.81...	1.93...	-0.4...	0.57...	0.501	-2.1...	0.595	1.25...	-0.3...	-0.8...	0.179	0.01...	0.77...	-4.5...	-0.0...	-0.5...	-0.5...	-0.0...	-0.			
3	0.06...	0.27...	0.47...	0.29...	0.99	-0.1...	0.20...	0.035	-0.5...	0.011	0.65...	-0.8...	2.06...	-0.3...	0.57...	-0.2...	-2.0...	0.193	0.31...	0.06...	0.12...	0.			
4	-0.1...	-0.8...	0.70...	0.23...	-1.7...	0.93...	-0.4...	0.57...	0.11...	-0.118	0.22...	-0.4...	2.532	-0.603	0.76...	-0.3...	-3.0...	0.565	0.554	0.16...	-0.2...	0.			
5	0.06...	-0.7...	0.62...	0.44...	0.70...	0.10...	-0.3...	0.47...	0.17...	-0.344	0.93...	-0.1...	1.186	-0.7...	0.56...	0.18...	-3.6...	0.09...	0.10...	-0.0...	-0.3...	0.			
6	0.28...	0.16...	0.07...	-0.0...	0.19...	-0.2...	-0.4...	0.23...	0.098	0.19	1.14...	-0.912	3.975	-0.9...	1.00...	0.20...	-2.7...	0.51...	-0.2...	-0.1...	5.41...	0.			
7	0.42...	0.143	0.92...	0.186	1.32...	1.15...	-0.4...	0.39...	-0.0...	-0.1...	0.60...	-1.0...	2.55...	-0.7...	0.38...	-0.0...	-2.2...	-0.0...	0.65...	0.07	0.92...	0.			
8	0.06...	-0.5...	0.50...	0.23...	0.05...	-0.114	-0.3...	0.34...	0.08...	-0.234	0.73...	-0.834	2.16...	-0.557	0.428	0.01...	-3.5...	0.23...	0.23...	0.02...	0.22...	0.			
9	0.23...	-0.0...	0.68...	0.24...	0.80...	1.71...	0.13...	0.44	0.86...	-0.1...	1.02...	-0.6...	1.579	-0.5...	0.32...	0.02...	-2.7...	0.10...	0.68...	-0.2...	-0.534	0.			
10	0.25...	-0.0...	0.63...	0.164	0.59...	0.14...	-0.1...	0.476	0.39...	0.10...	1.12...	-0.523	2.00...	-0.8...	0.48...	0.24...	-2.9...	0.51...	0.36...	-0.0...	0.21...	0.			
11	0.47...	-0.0...	0.12...	1.26...	0.64...	0.18...	0.55	0.48...	-2.0...	0.33	2.17...	-0.2...	-1.1...	-0.0...	-0.1...	1.35...	-2.3...	-0.8...	-0.3...	-0.3...	0.17...	0.			
12	0.20...	-0.4...	0.91...	0.00...	0.05...	0.79...	-0.435	0.59	-0.4...	0.123	0.70...	-0.543	2.328	-0.8...	0.358	0.016	-2.7...	0.30...	0.64...	-0.0...	0.00...	0.			
13	0.11...	0.38...	0.80...	-0.051	0.10...	0.67...	-0.3...	0.35...	0.25...	0.01...	0.18...	-1.0...	3.08...	-0.613	0.58...	0.11...	-3.2...	0.72...	0.58...	-0.0...	0.14...	0.			
14	-0.5...	0.53...	1.756	0.25...	-0.4...	-1.2...	0.14...	-0.3...	0.06...	0.71...	1.41...	-0.609	1.943	-0.1...	-0.5...	0.27...	-2.258	-0.1...	1.59...	-0.727	0.47...	0.			
15	-1.7...	0.487	1.85...	-0.0...	0.48...	0.133	-0.3...	0.38...	-0.4...	-0.1...	1.24...	-1.0...	1.75...	0.05...	0.38...	0.39...	-2.6...	-0.5...	1.71...	-0.6...	0.96	-0.			
16	-1.331	0.08...	1.53...	0.25...	-0.6...	-0.2...	-0.1...	0.30...	-0.4...	-0.05	1.463	-0.7...	2.01...	0.288	0.23...	0.75...	-2.5...	0.35...	1.28...	-1.0...	0.07...	-0.			
17	-0.633	0.09...	0.75...	0.82...	0.67...	0.84	0.37...	0.69...	-0.6...	0.14...	1.60...	-0.4...	1.35...	-0.2...	0.07...	0.55...	-2.9...	0.10...	0.45...	-0.5...	1.49...	0.			
18	0.558	0.29...	0.01...	0.82...	0.69...	-0.1...	0.19...	0.213	0.29...	-0.182	1.14...	-0.5...	2.82...	-0.9...	0.92...	-0.3...	-2.013	-0.0...	-0.5...	0.46...	0.09...	0.			
19	-0.2...	-0.599	1.20...	-0.0...	-0.8...	1.407	-0.5...	0.34...	0.19...	-0.427	0.19...	-1.0...	2.37...	-0.624	0.66...	-0.4...	-3.6...	0.17...	1.19...	0.05...	0.35...	0.			
20	0.00...	-0.4...	0.49...	1.459	-0.3...	0.29...	0.46...	0.61...	-0.5...	-0.134	2.24...	-1.0...	-0.5...	-0.1...	0.13...	1.164	-1.437	0.19...	-0.4...	-0.4...	-0.1...	0.			
21	0.00...	-0.1...	0.18...	0.70...	1.32...	0.36...	0.82...	-2.0...	0.20...	2.35...	-0.2...	-1.56	-0.2...	0.46...	1.53...	-4.0...	-1.0...	-0.4...	-0.3...	0.97...	0.				
22	-0.2...	0.32...	0.92...	0.233	0.01...	1.15	-0.2...	0.268	-0.0...	0.17...	0.80...	-0.959	1.60...	-0.6...	0.79...	0.67...	-3.1...	-0.0...	0.97...	-0.1...	-0.2...	0.			
23	-0.5...	0.039	0.99...	0.34...	-1.1...	-0.816	-0.3...	0.177	-0.0...	-0.0...	1.339	-0.3...	2.533	-1.0...	0.78...	0.08...	-3.3...	0.85...	0.93...	-0.3...	0.43...	0.			
24	-0.5...	-0.5...	1.60...	-0.2...	0.177	0.05...	-0.4...	0.72...	-0.1...	0.608	-0.016	-0.4...	1.94...	-0.4...	0.015	-0.1...	-4.5...	0.67...	1.49...	-0.2...	1.02...	0.			
25	0.07...	-0.0...	0.66...	0.22...	1.48...	0.289	-0.128	0.526	-0.1...	0.06...	1.142	-0.155	1.89...	-0.7...	0.44...	0.29...	-2.8...	0.01...	-9.2...	-0.277	0.12...	0.			
26	-0.5...	0.06...	0.42...	-0.0...	0.213	0.62...	-0.4...	0.76...	0.49...	0.58...	0.53...	1.24...	2.439	-1.3...	0.08...	-0.0...	-4.3...	-0.1...	0.10...	-0.1...	0.28...	0.			
27	-0.5...	-0.1...	0.92...	0.02...	0.12...	0.23...	-0.1...	0.328	-0.0...	0.05...	0.586	0.26...	2.52...	-0.7...	-0.0...	0.00...	-2.9...	0.31...	1.23...	-0.3...	0.34...	-0.			
28	0.27...	-0.1...	0.60...	-0.2...	0.38	0.69...	-0.4...	0.46...	0.592	0.15...	0.64...	-0.2...	2.22...	-1.3...	0.17...	-0.3...	-4.0...	0.19...	0.24...	0.07...	-0.1...	0.			
29	0.127	0.331	0.36...	-0.2...	0.42...	0.00...	-0.6...	0.48...	-0.2...	0.481	-0.0...	0.42...	1.77...	-0.2...	-0.0...	-0.2...	-4.1...	-0.1...	-0.0...	-0.0...	0.615	0.			
30	-0.1...	-0.2...	0.63...	0.87...	0.268	0.32...	0.06...	0.23...	-0.5...	-0.0...	1.604	-0.8...	1.44...	0.12...	0.74...	0.28	-2.3...	-0.2...	0.06...	-0.2...	-0.5...	0.			
31	0.21...	0.40...	0.26...	1.071	1.50...	-0.1...	0.33...	0.62...	-2.6...	0.493	1.85...	-0.433	0.36...	-0.185	0.32...	1.48...	-3.3...	0.57...	-0.2...	-0.4...	-0.069	0.			
32	-0.264	-0.4...	0.64...	0.86...	-0.0...	-0.0...	0.231	0.90...	-0.1...	-0.095	1.58...	-0.9...	1.19...	-0.0...	0.15...	1.78...	-3.0...	0.19...	0.30...	-0.2...	0.52...	0.			
33	-0.2...	-0.3...	1.42...	-0.1...	0.16...	1.37...	-0.4...	0.476	-0.0...	0.338	0.25...	-0.6...	1.663	-0.3...	0.164	-0.1...	-2.7...	0.06...	1.10...	-0.0...	0.56...	0.			
34	-0.3...	-0.194	1.14...	0.22...	0.39...	-0.441	-0.2...	0.27...	-0.6...	0.102	0.85...	-0.2...	2.31...	-0.4...	0.221	-0.1...	-4.2...	0.31...	0.70...	-0.1...	0.56...	0.			

Convert Dataset from CSV to ARFF format



Weka Explorer - Data Load



Current relation

Relation: Combined
Instances: 590

Attributes: 17815
Sum of weights: 590

Attributes

All

None

Invert

Pattern

No.	Name
17798	17798
17799	17799
17800	17800
17801	17801
17802	17802
17803	17803
17804	17804
17805	17805
17806	17806
17807	17807
17808	17808
17809	17809
17810	17810
17811	17811
17812	17812
17813	17813
17814	17814
17815	label

Remove

Selected attribute

Name: label
Missing: 0 (0%)

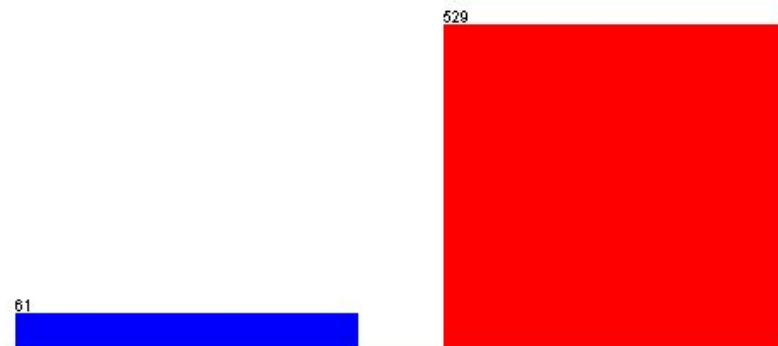
Distinct: 2

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	normal	61	61.0
2	tumor	529	529.0

Class: label (Nom)

Visualize All



Status

Current relation

Relation: Combined
Instances: 590

Attributes: 17815
Sum of weights: 590

Attributes

All

None

Invert

Pattern

No.		Name
17586	<input type="checkbox"/>	17586
17587	<input type="checkbox"/>	17587
17588	<input type="checkbox"/>	17588
17589	<input type="checkbox"/>	17589
17590	<input type="checkbox"/>	17590
17591	<input type="checkbox"/>	17591
17592	<input type="checkbox"/>	17592
17593	<input checked="" type="checkbox"/>	17593
17594	<input type="checkbox"/>	17594
17595	<input type="checkbox"/>	17595
17596	<input type="checkbox"/>	17596
17597	<input type="checkbox"/>	17597
17598	<input type="checkbox"/>	17598
17599	<input type="checkbox"/>	17599
17600	<input type="checkbox"/>	17600
17601	<input type="checkbox"/>	17601
17602	<input type="checkbox"/>	17602
17603	<input type="checkbox"/>	17603

Remove

Selected attribute

Name: 17593
Missing: 0 (0%)

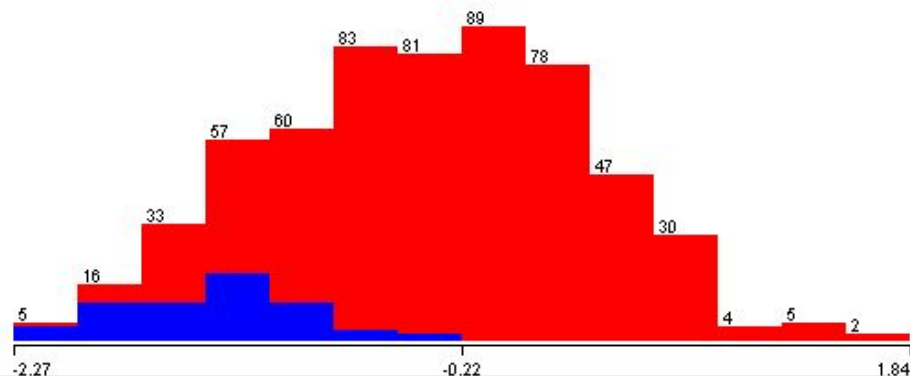
Distinct: 577

Type: Numeric
Unique: 564 (96%)

Statistic	Value
Minimum	-2.272
Maximum	1.835
Mean	-0.387
StdDev	0.721

Class: label (Nom)

Visualize All



SVM - SMO

[illegible]

```
weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8
```

```
to build model: 1.57 seconds
```

Implements John Platt's sequential minimal optimization algorithm for training a support vector classifier

This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default. (In that case the coefficients in the output are based on the normalized data, not the original data — this is important for interpreting the classifier.)

Multi-class problems are solved using pairwise classification (aka 1-vs-1).

To obtain proper probability estimates, use the option that fits calibration models to the outputs of the support vector machine. In the multi-class case, the predicted probabilities are coupled using Hastie and Tibshirani's pairwise coupling method.

Note: for improved speed normalization should be turned off when operating on SparseInstances.

For more information on the SMO algorithm, see

J. Platt: Fast Training of Support Vector Machines using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, 1998.

S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13(3):637-649.

Trevor Hastie, Robert Tibshirani: Classification by Pairwise Coupling. In: Advances in Neural Information Processing Systems, 1998.

CAPABILITIES

Class – Binary class, Missing class values, Nominal class

Attributes – Binary attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces – WeightedInstancesHandler

Additional

Minimum number of instances: 1

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic

Test options

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

(Nom) label

Result list (right-click for options)

18:31:53 - functions.SMO

Classifier output

Time taken to build model: 0.75 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	588	99.661 %
Incorrectly Classified Instances	2	0.339 %
Kappa statistic	0.982	
Mean absolute error	0.0034	
Root mean squared error	0.0582	
Relative absolute error	1.8168 %	
Root relative squared error	19.1218 %	
Total Number of Instances	590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.004	0.968	1.000	0.984	0.982	0.998	0.968	normal
	0.996	0.000	1.000	0.996	0.998	0.982	0.998	1.000	tumor
Weighted Avg.	0.997	0.000	0.997	0.997	0.997	0.982	0.998	0.996	

=== Confusion Matrix ===

```

a  b  <-- classified as
61  0 |  a = normal
 2 527 | b = tumor

```

X: False Positive Rate (Num) ▾

Y: True Positive Rate (Num) ▾

Colour: Threshold (Num) ▾

Select Instance ▾

Reset

Clear

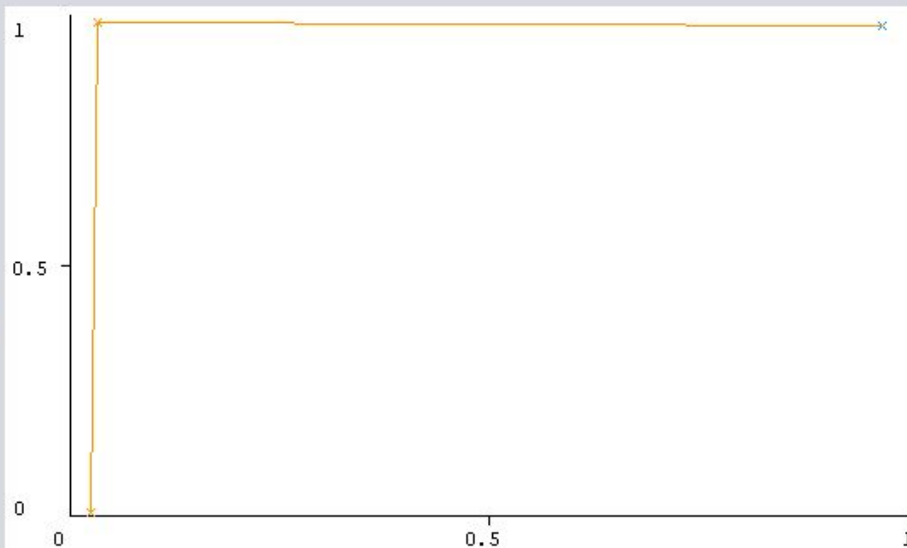
Open

Save

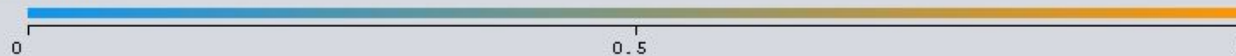
Jitter



Plot (Area under ROC = 0.9981)



Class colour



X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Reset

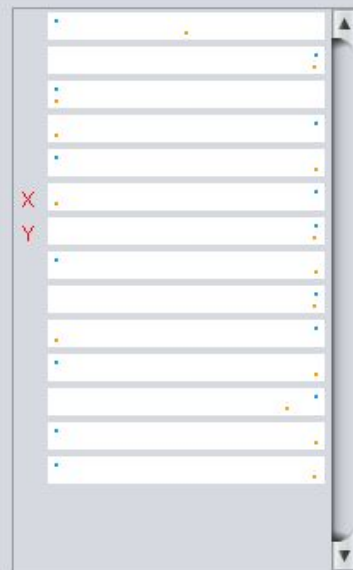
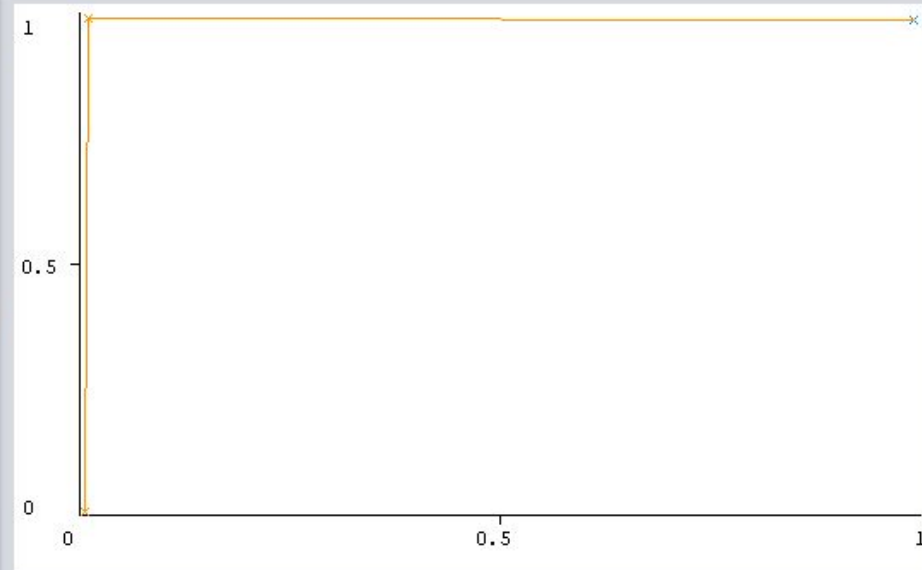
Clear

Open

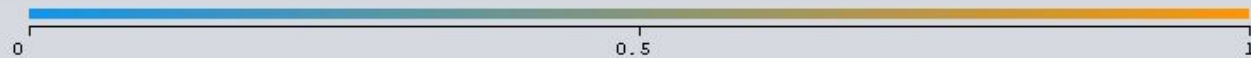
Save

Jitter

Plot (Area under ROC = 0.9981)



Class colour



Decision Tree

“Rep Tree”

Reduced Error Pruning Tree (RepTree) is a fast decision tree learner that builds a decision/regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning algorithm.

Classifier

- ▼ weka
 - ▼ classifiers
 - bayes
 - functions
 - lazy
 - meta
 - misc
 - rules
 - ▼ trees
 - DecisionStump
 - HoeffdingTree
 - J48
 - LMT
 - M5P
 - RandomForest
 - RandomTree
 - REPTree**

Fast decision tree learner

Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting). Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5).

CAPABILITIES

Class -- Binary class, Date class, Missing class values, Nominal class, Numeric class

Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces -- Drawable, PartitionGenerator, Randomizable, Sourcable, WeightedInstancesHandler

Additional

Minimum number of instances: 1

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

Set...

☒ Cross-validation Folds

☐ Percentage split %

More options...

(Nom) label

Start

Stop

Result list (right-click for options)

18:22:30 - trees.REPTree

Classifier output

Size of the tree : 3

Time taken to build model: 2.44 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	578	97.9661 %
Incorrectly Classified Instances	12	2.0339 %
Kappa statistic	0.8919	
Mean absolute error	0.0298	
Root mean squared error	0.1383	
Relative absolute error	15.9835 %	
Root relative squared error	45.4205 %	
Total Number of Instances	590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.918	0.013	0.889	0.918	0.903	0.892	0.937	0.846	normal
	0.987	0.082	0.991	0.987	0.989	0.892	0.937	0.986	tumor
Weighted Avg.	0.980	0.075	0.980	0.980	0.980	0.892	0.937	0.971	

=== Confusion Matrix ===

```

a  b  <-- classified as
56  5 |  a = normal
7 522 |  b = tumor

```

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Reset

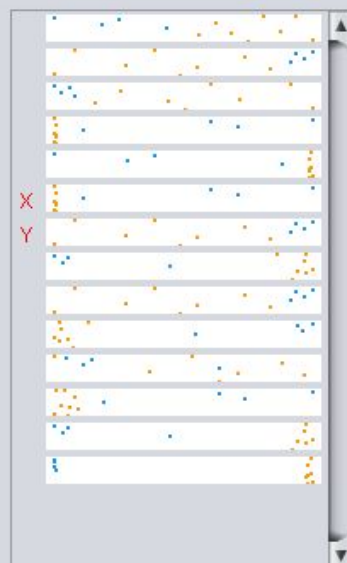
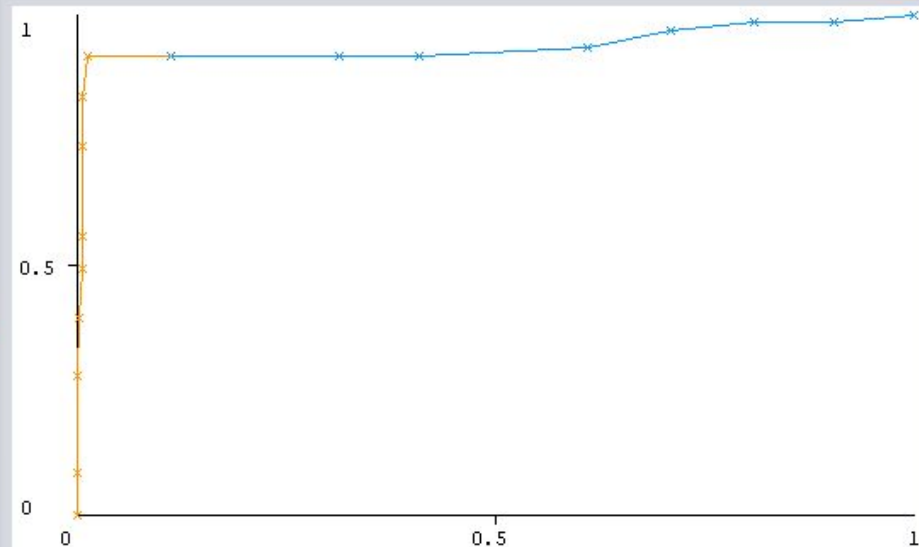
Clear

Open

Save

Jitter ☐

Plot (Area under ROC = 0.9371)



Class colour

0.0021

0.49

0.98

X: False Positive Rate (Num) ▼

Y: True Positive Rate (Num) ▼

Colour: Threshold (Num) ▼

Select Instance ▼

Reset

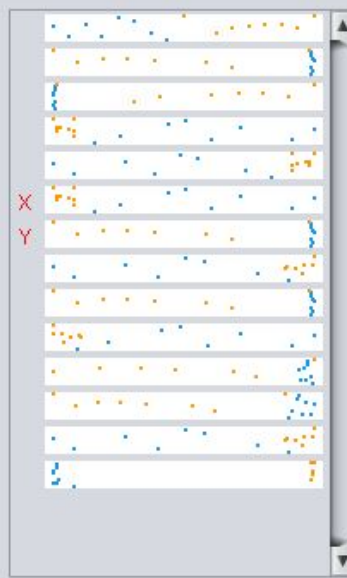
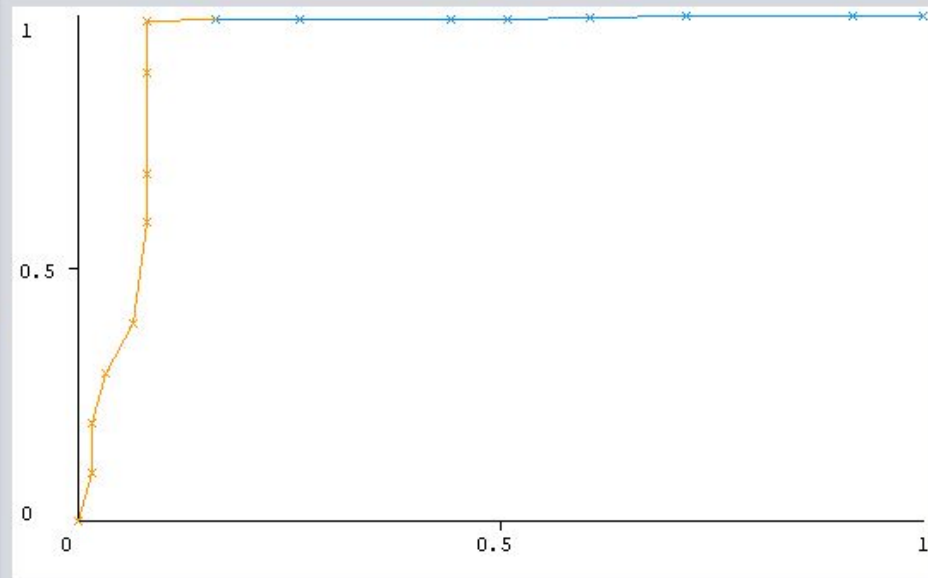
Clear

Open

Save

Jitter ☐

Plot (Area under ROC = 0.9371)



Class colour

0.019

0.51

1

Random Forest

Classifier

- ▼ weka
 - ▼ classifiers
 - bayes
 - functions
 - lazy
 - meta
 - misc
 - rules
 - ▼ trees
 - DecisionStump
 - HoeffdingTree
 - J48
 - LMT
 - M5P
 - RandomForest**
 - RandomTree
 - RE

.0 -V 0.001 -S 1

put

Class for constructing a forest of random trees

For more information see:

Leo Breiman (2001). Random Forests. Machine Learning. 45(1):5-32.

CAPABILITIES

Class -- Binary class, Missing class values, Nominal class, Numeric class

Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces -- Drawable, PartitionGenerator, Randomizable, WeightedInstancesHandler

Additional

Minimum number of instances: 1

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
- ☐ Percentage split %

(Nom) label ▼

Result list (right-click for options)

18:16:49 - trees.RandomForest

Classifier output

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

Time taken to build model: 2.22 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	585	99.1525 %
Incorrectly Classified Instances	5	0.8475 %
Kappa statistic	0.9533	
Mean absolute error	0.0458	
Root mean squared error	0.1041	
Relative absolute error	24.5672 %	
Root relative squared error	34.1805 %	
Total Number of Instances	590	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.934	0.002	0.983	0.934	0.958	0.954	0.999	0.995	normal
	0.998	0.066	0.992	0.998	0.995	0.954	0.999	1.000	tumor
Weighted Avg.	0.992	0.059	0.991	0.992	0.991	0.954	0.999	0.999	

=== Confusion Matrix ===

```
a  b  <-- classified as
57  4  |  a = normal
1 528 |  b = tumor
```

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Reset

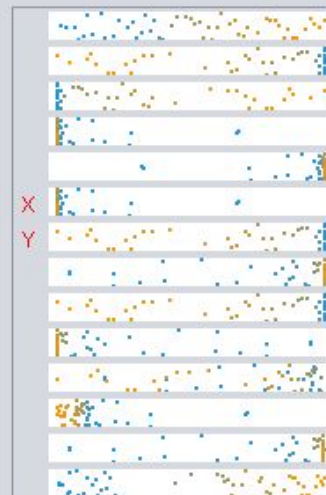
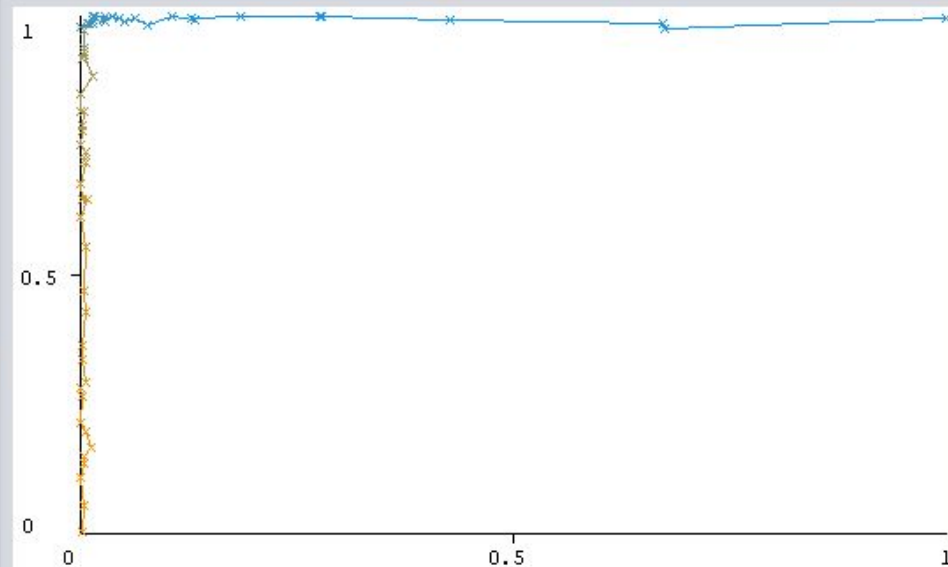
Clear

Open

Save

Jitter

Plot (Area under ROC = 0.9995)



Class colour

0

0.48

0.95

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Reset

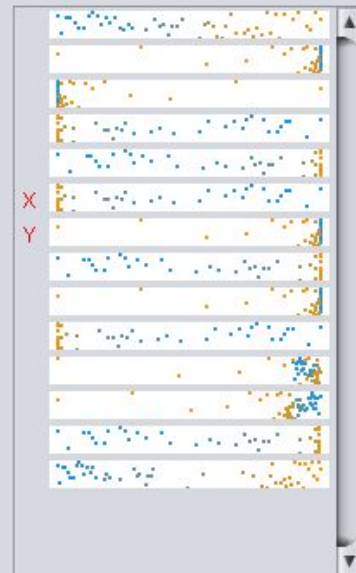
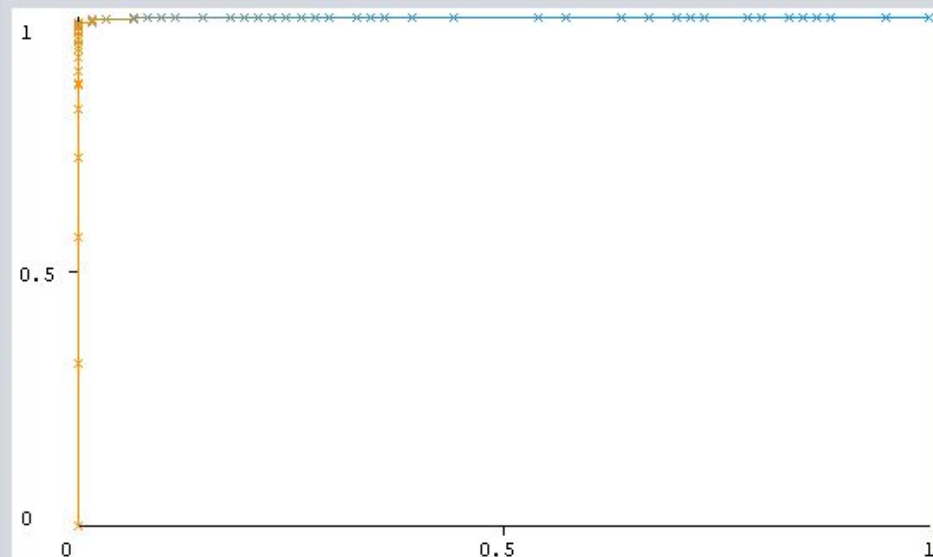
Clear

Open

Save

Jitter

Plot (Area under ROC = 0.9995)



Class colour

