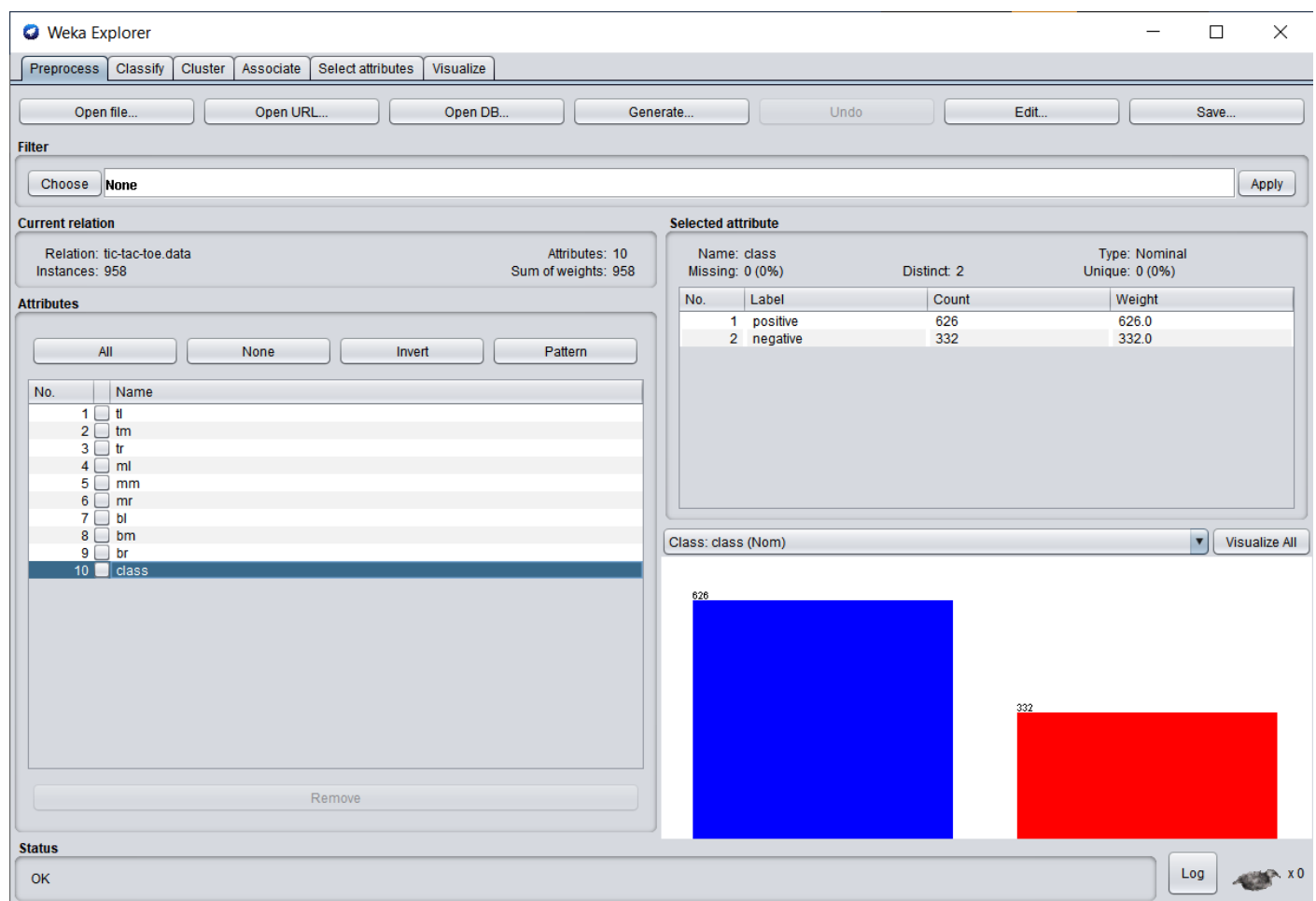


LAB – 02

Illustrate a classification algorithm, a clustering algorithm and a regression algorithm using Weka Tool. Evaluate suitable performance metrics.

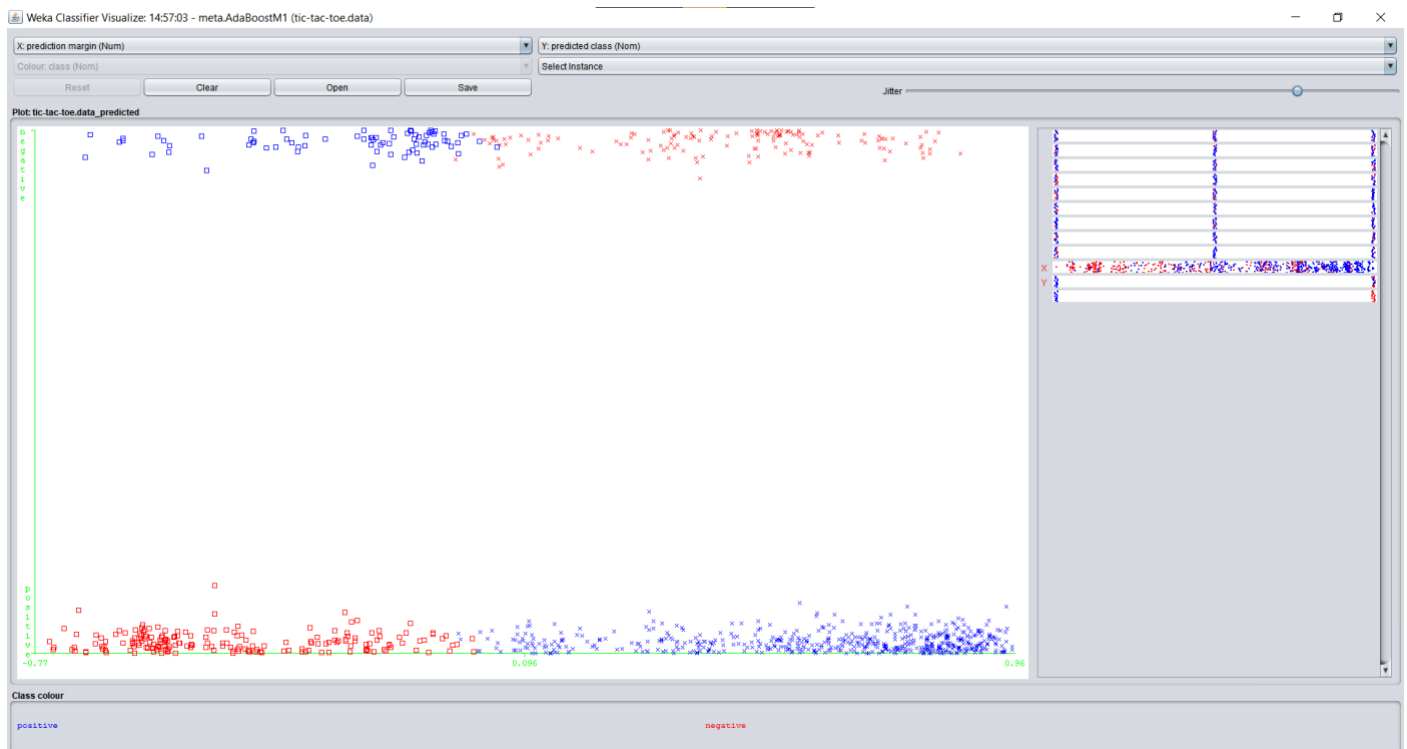
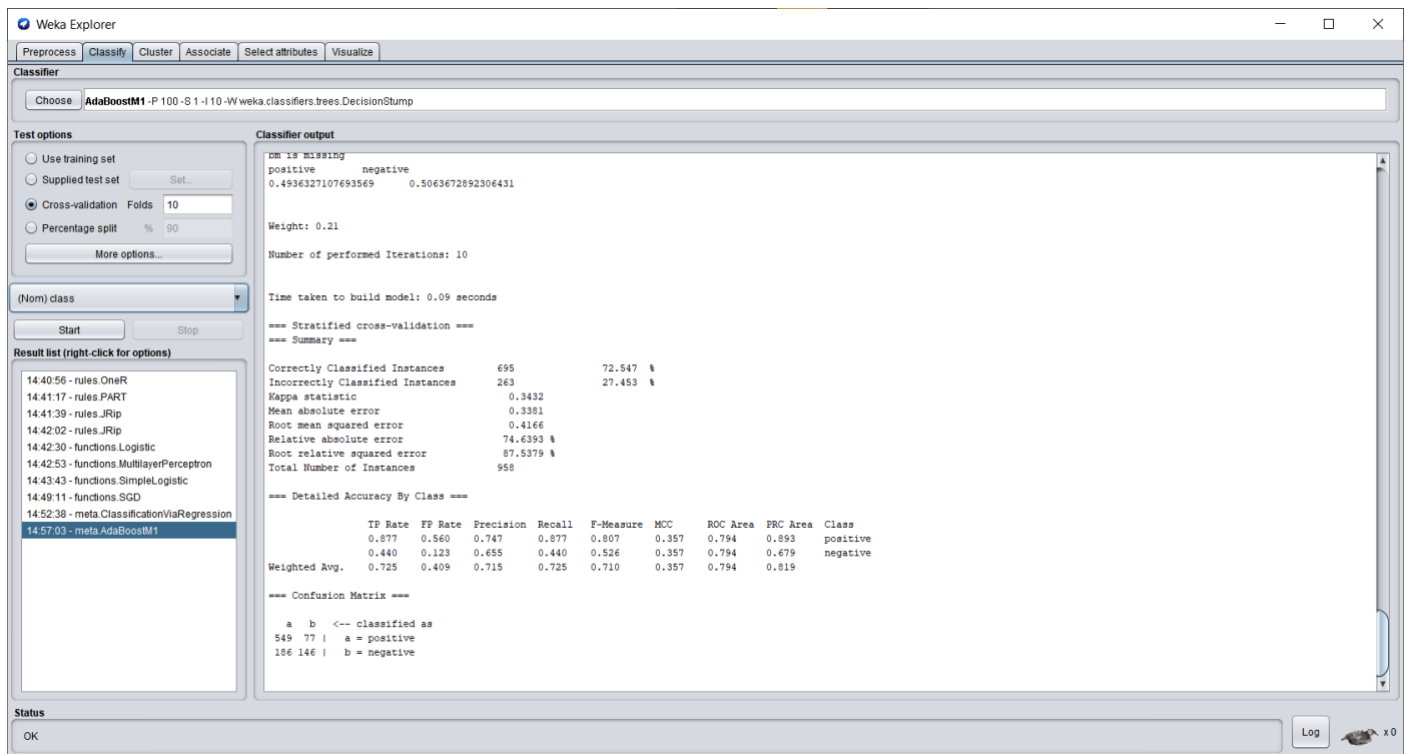
Dataset:-

[Index of /ml/machine-learning-databases/tic-tac-toe \(uci.edu\)](http://index.of/ml/machine-learning-databases/tic-tac-toe(uci.edu))

Preprocess:-

Algorithms:-

1. AdaBoostM1 (Classifier)



The model was quick to compile but very bad at the actual classification, as can be viewed from the confusion matrix.

2. Logistic (Regression)

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **AdaBoostM1 - P 100 - S 1 - I 10 - W weka.classifiers.trees.DecisionStump**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 90

More options...

(Nom) class

Start Stop

Result list (right-click for options)

- 14:40:56 - rules.OneR
- 14:41:17 - rules.PART
- 14:41:39 - rules.JRip
- 14:42:02 - rules.JRip
- 14:42:30 - functions.Logistic**
- 14:42:53 - functions.MultilayerPerceptron
- 14:43:43 - functions.SimpleLogistic
- 14:49:11 - functions.SGD
- 14:52:38 - meta.ClassificationViaRegression
- 14:57:03 - meta.AdaBoostM1

Classifier output

```
LM=x 2.300352200020504e19
bm=b 77.1965
br=o 0
br=x 1.2640131674498609e19
br=b 387.2291

Time taken to build model: 0.29 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 942 98.3299 %
Incorrectly Classified Instances 16 1.6701 %
Kappa statistic 0.9627
Mean absolute error 0.029
Root mean squared error 0.1257
Relative absolute error 6.3994 %
Root relative squared error 26.4181 %
Total Number of Instances 958

=== Detailed Accuracy By Class ===

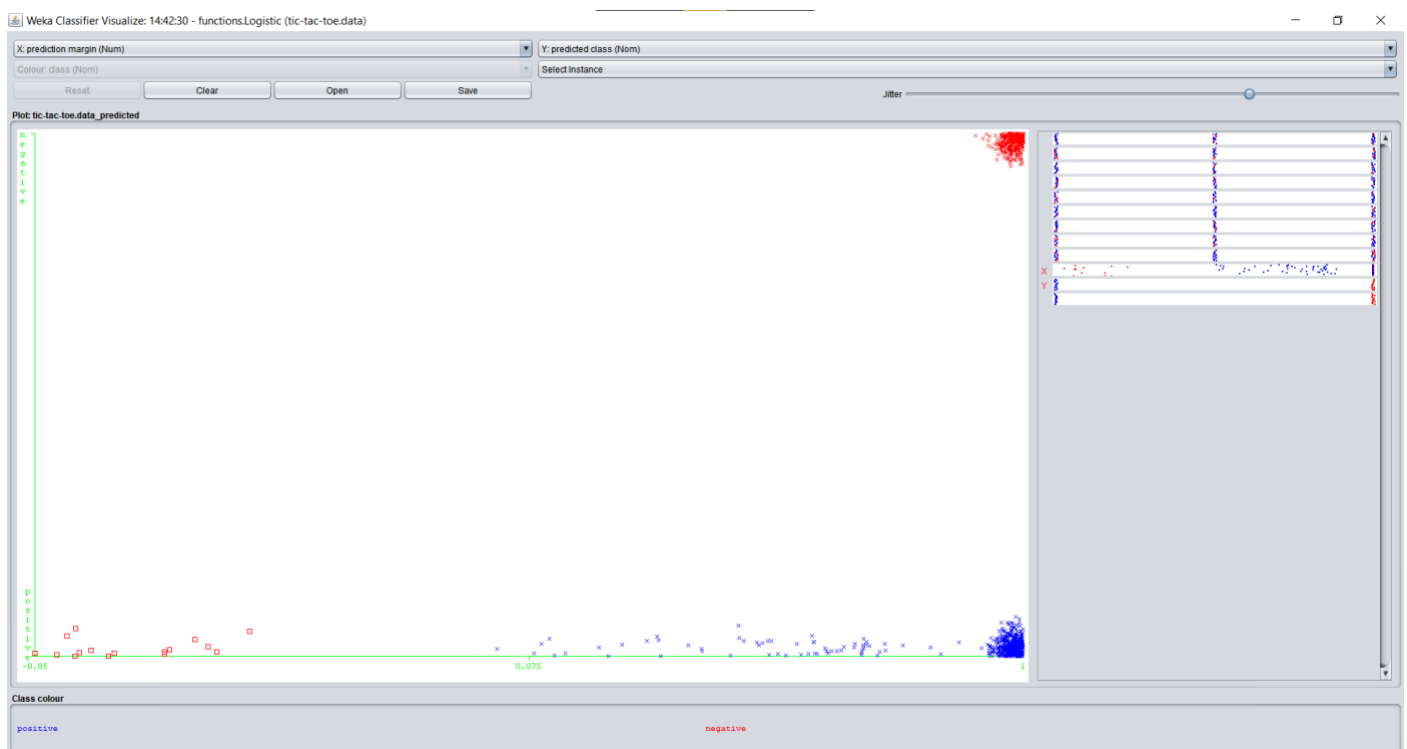
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          1.000    0.048    0.975    1.000    0.987    0.963    0.997    0.998    positive
          0.952    0.000    1.000    0.952    0.975    0.963    0.997    0.994    negative
Weighted Avg.  0.983    0.031    0.984    0.983    0.983    0.963    0.997    0.997

=== Confusion Matrix ===

  a  b  <-- classified as
626  0 | a = positive
 16 316 | b = negative
```

Status

OK Log x 0



3. Stochastic Gradient Descent

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose SGD-F 0-L 0.01-R 1.0E-4-E 500-C 0.001-S 1

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 90
 More options...

(Nom) class

Start Stop

Result list (right-click for options)

- 14:40:56 - rules.OneR
- 14:41:17 - rules.PART
- 14:41:39 - rules.JRip
- 14:42:02 - rules.JRip
- 14:42:30 - functions.Logistic
- 14:42:53 - functions.MultilayerPerceptron
- 14:43:43 - functions.SimpleLogistic
- 14:49:11 - functions.SGD**

Classifier output

```

+ 2.4498 (normalized) bl=a
+ 0.06 (normalized) bl=b
+ 2.4498 (normalized) bm=a
+ -2.2289 (normalized) bm=x
+ 0.06 (normalized) bm=b
+ 2.4288 (normalized) br=a
+ -2.2189 (normalized) br=x
+ 0.07 (normalized) br=b
+ 0.28

Time taken to build model: 0.93 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      942      98.3299 %
Incorrectly Classified Instances    16       1.6701 %
Kappa statistic                    0.9627
Mean absolute error                 0.0167
Root mean squared error             0.1292
Relative absolute error             3.6967 %
Root relative squared error        27.1571 %
Total Number of Instances          958

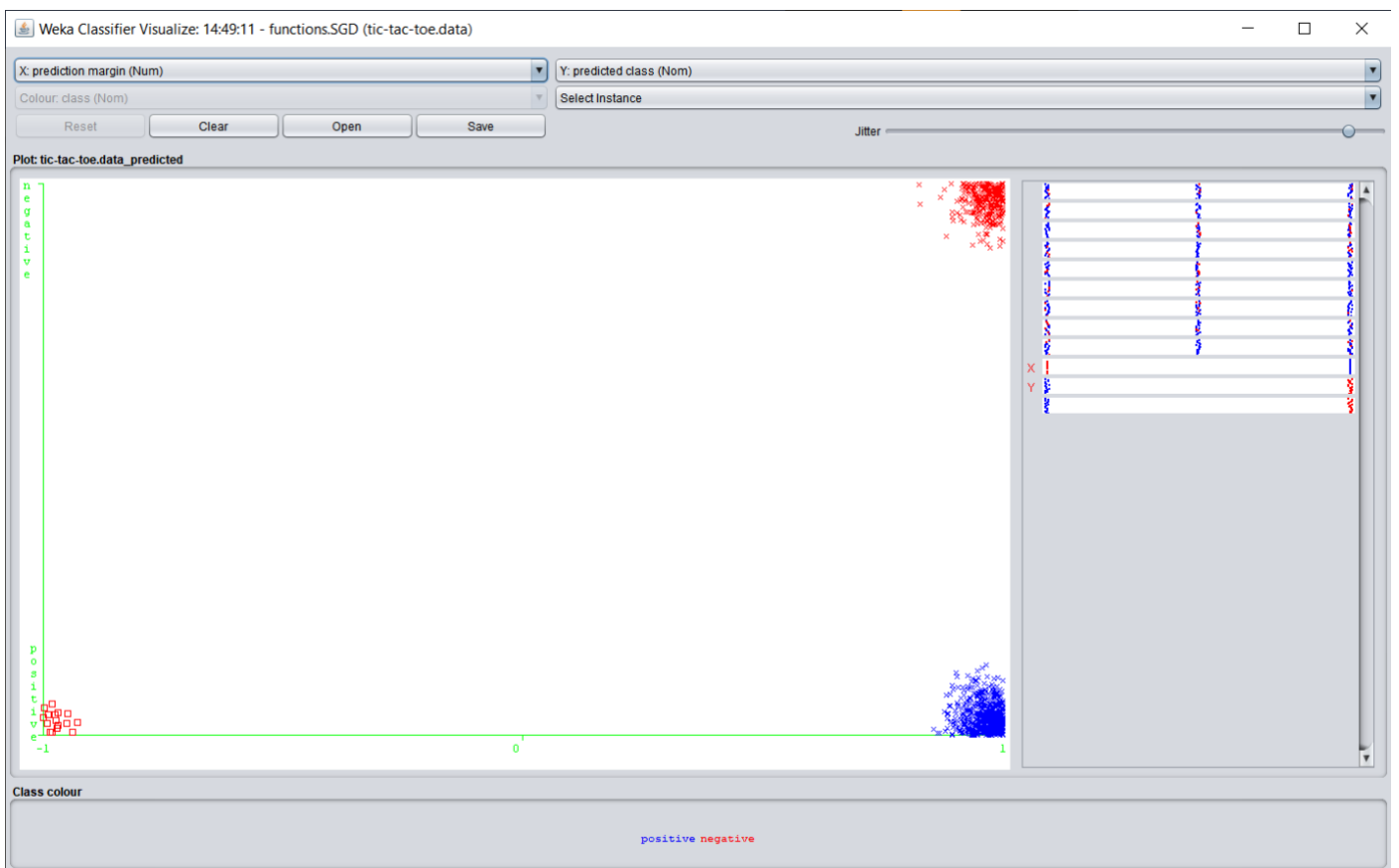
=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
    1.000    0.048    0.975    1.000    0.987    0.963    0.976    0.975    positive
    0.952    0.000    1.000    0.952    0.975    0.963    0.976    0.969    negative
Weighted Avg.   0.983    0.031    0.984    0.983    0.983    0.963    0.976    0.973

=== Confusion Matrix ===
  a  b  <-- classified as
626  0 | a = positive
 16 316 | b = negative
  
```

Status

OK Log



4. Classification via Regression

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ClassificationViaRegression -VV weka.classifiers.trees.M5P -- -M 4.0

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 90

More options...

(Nom) class

Start Stop

Result list (right-click for options)

- 14.40.56 - rules OneR
- 14.41.17 - rules PART
- 14.41.39 - rules JRip
- 14.42.02 - rules JRip
- 14.42.30 - functions Logistic
- 14.42.53 - functions MultilayerPerceptron
- 14.43.43 - functions SimpleLogistic
- 14.49.11 - functions SGD
- 14.52.38 - meta ClassificationViaRegression**

Classifier output

```
- 0.5914 * cm=b,o,x
- 0.9291 * km=x
+ 0.9326 * br=b,o
+ 0.9939 * bz=b
+ 0.0746
```

Number of Rules : 1

Time taken to build model: 1.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	899	93.8413 %
Incorrectly Classified Instances	59	6.1587 %
Kappa statistic	0.8624	
Mean absolute error	0.1363	
Root mean squared error	0.2269	
Relative absolute error	30.0972 %	
Root relative squared error	47.6665 %	
Total Number of Instances	958	

=== Detailed Accuracy By Class ===

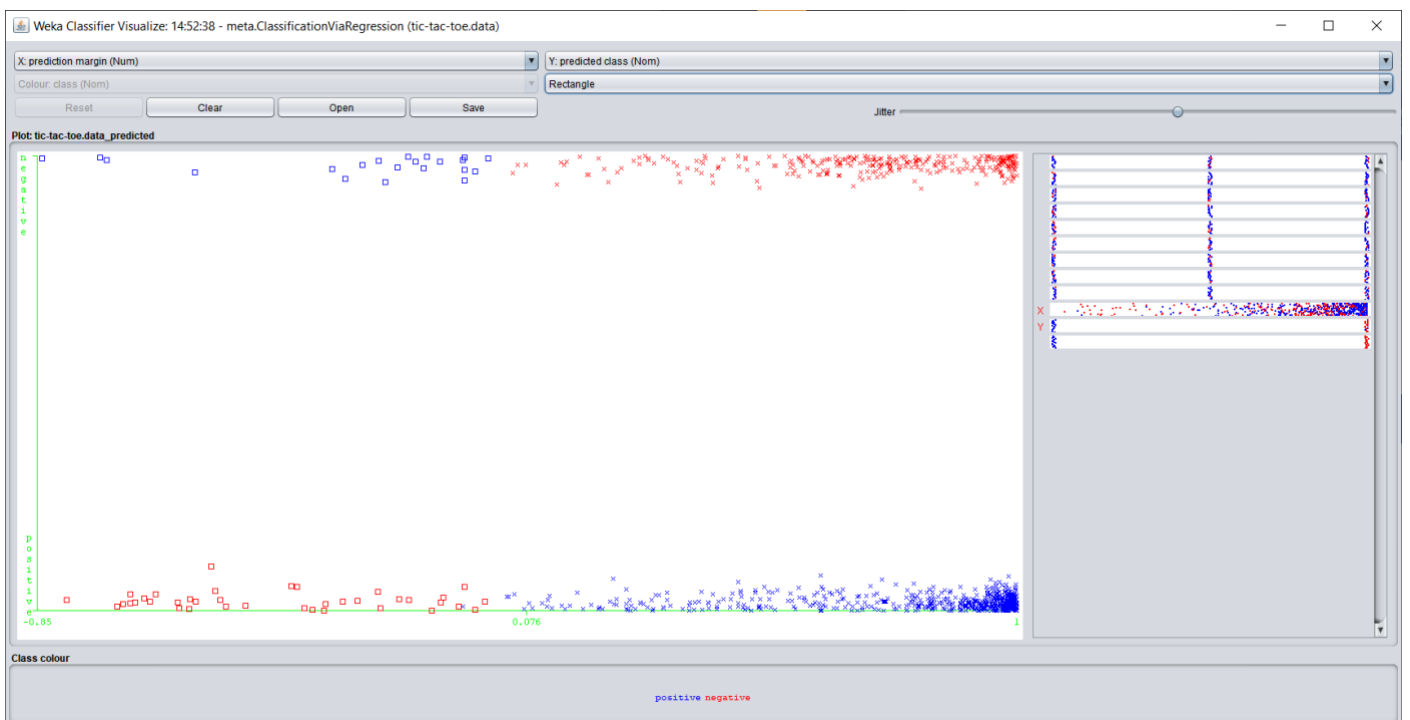
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.566	0.114	0.941	0.966	0.554	0.863	0.982	0.990	positive
	0.886	0.034	0.933	0.886	0.909	0.863	0.982	0.970	negative
Weighted Avg.	0.938	0.056	0.938	0.938	0.938	0.863	0.982	0.983	

=== Confusion Matrix ===

a	b	classified as	
605	21		a = positive
38	294		b = negative

Status

OK Log



Clustering:-

1. Simple K-Means

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -ft -1.25 -t2 -1.0 -N 2 -A *weka.core.EuclideanDistance -R first-last -i 500 -num-slots 1 -S 10

Cluster mode

☐ Use training set
☐ Supplied test set
☐ Percentage split % 66
☒ Classes to clusters evaluation
(Nom) class
☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

- 15:14:13 - SimpleKMeans
- 15:14:43 - HierarchicalClusterer
- 15:15:32 - FarthestFirst
- 15:15:40 - Canopy
- 15:16:06 - Cobweb
- 15:17:05 - SimpleKMeans

Clusterer output

Cluster 0: x,x,x,b,b,b,b,b,b,b
Cluster 1: x,x,x,b,b,b,b,b,b,b

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#
	(958.0)	(540.0) (418.0)
tl	x	x 0
tm	x	x x
tr	x	x x
ml	x	x 0
mm	x	0 x
mr	x	0 x
bl	x	0 x
bm	x	0 x
br	x	x 0

Time taken to build model (full training data) : 0.07 seconds

==> Model and evaluation on training set ==>

Clustered Instances

Cluster	Count	Percentage
0	540	(56%)
1	418	(44%)

Class attribute: class
Classes to Clusters:

Cluster	Assigned to Cluster	Count	Percentage
0	1 <-- assigned to cluster	341	285 positive
1	0 <-- assigned to cluster	199	133 negative

Cluster 0 <-- negative
Cluster 1 <-- positive

Incorrectly clustered instances : 474.0 49.4761 %

Status
OK

Log

2. Farthest First

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose FarthestFirst -N 2 -S 1

Cluster mode

☐ Use training set
☐ Supplied test set
☐ Percentage split % 66
☒ Classes to clusters evaluation
(Nom) class
☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

- 15:14:13 - SimpleKMeans
- 15:14:43 - HierarchicalClusterer
- 15:15:32 - FarthestFirst
- 15:15:40 - Canopy
- 15:16:06 - Cobweb
- 15:17:05 - SimpleKMeans
- 15:21:17 - HierarchicalClusterer
- 15:22:29 - HierarchicalClusterer
- 15:23:11 - FarthestFirst

Clusterer output

tl
tm
tr
ml
mm
mr
bl
bm
br

Ignored: class

Test mode: Classes to clusters evaluation on training data

==> Clustering model (full training set) ==>

FarthestFirst

Cluster centroids:

Cluster 0
x x b o x o x b

Cluster 1
o o x x o x b b x

Time taken to build model (full training data) : 0.09 seconds

==> Model and evaluation on training set ==>

Clustered Instances

Cluster	Count	Percentage
0	550	(58%)
1	400	(42%)

Class attribute: class
Classes to Clusters:

Cluster	Assigned to Cluster	Count	Percentage
0	1 <-- assigned to cluster	380	246 positive
1	0 <-- assigned to cluster	178	154 negative

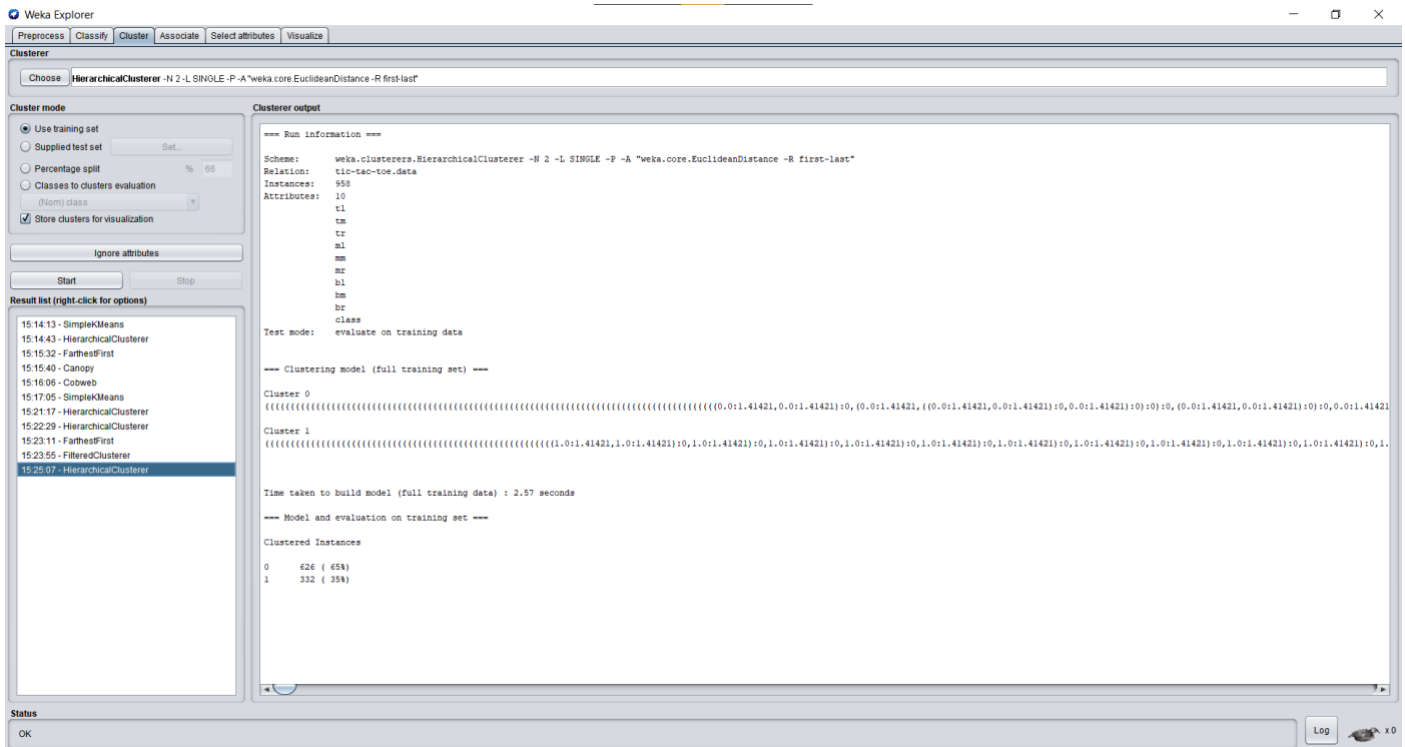
Cluster 0 <-- positive
Cluster 1 <-- negative

Incorrectly clustered instances : 424.0 44.2559 %

Status
OK

Log

3. Hierarchical Cluster



4. EM

