

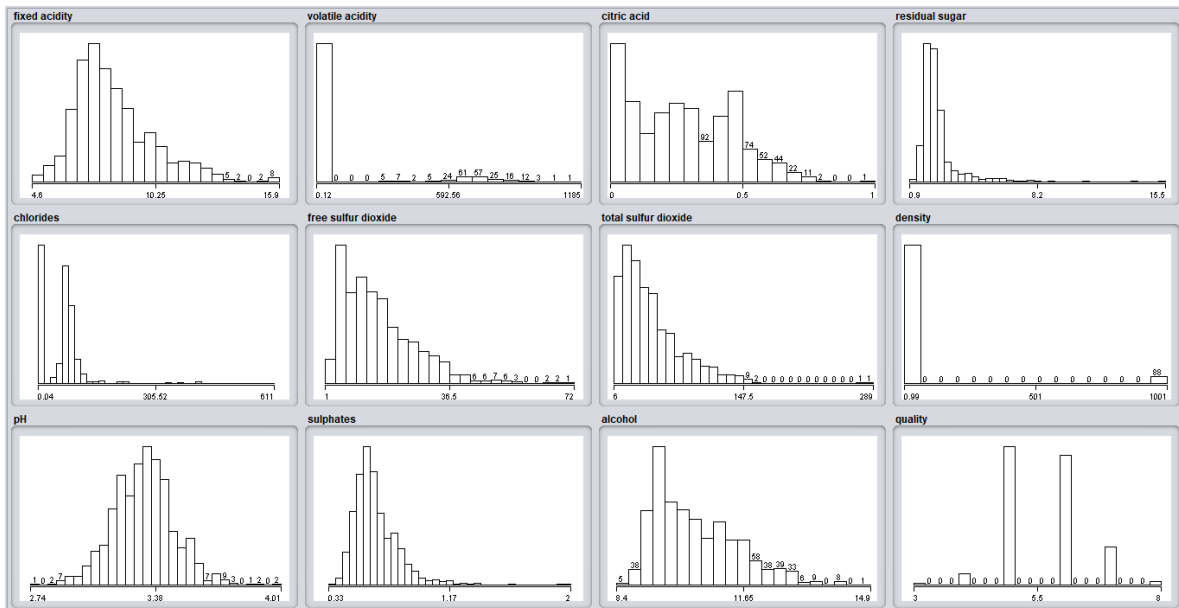
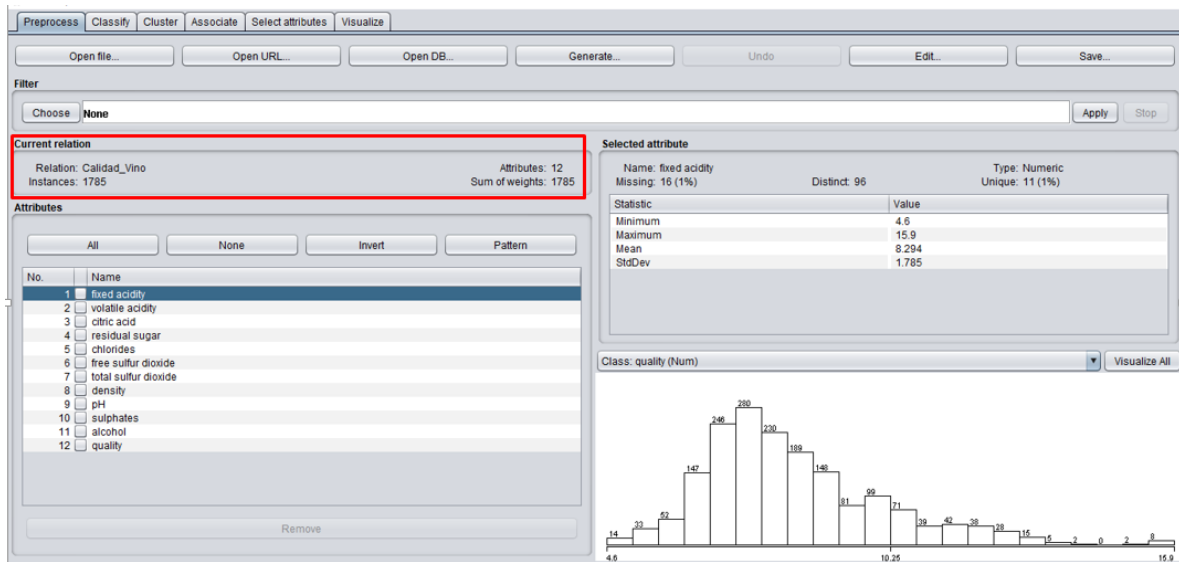
Laura Cristina López Bedoya

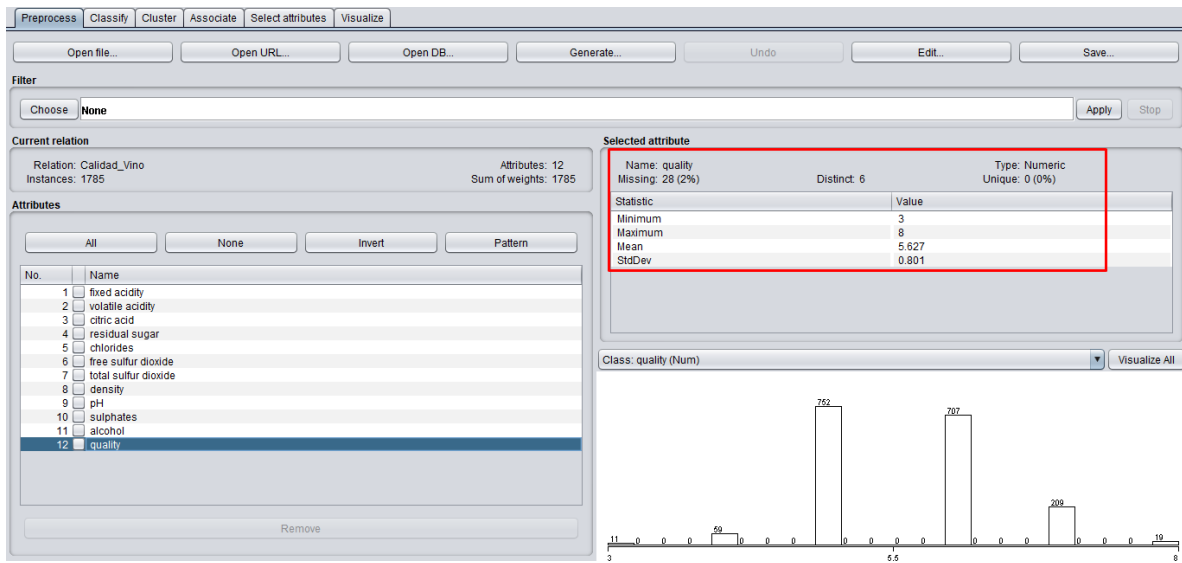
WEKA

PREPARACIÓN DE DATOS

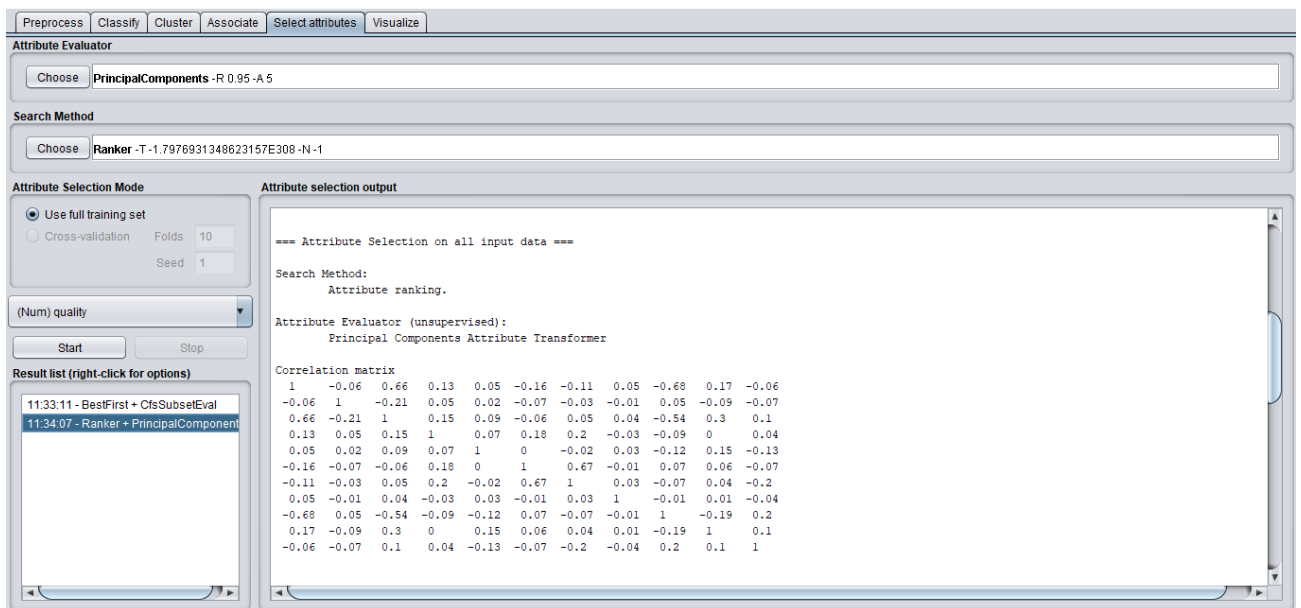
Al cargar el archivo Calidad_Vino.arff en Weka se inicia con la preparación y visualización de datos en el programa.

- Visualización de los datos:





- Matriz de correlación:



Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **PrincipalComponents -R 0.95-A 5**

Search Method

Choose **Ranker -T-1.7976931348623157E308-N-1**

Attribute Selection Mode

☒ Use full training set
☐ Cross-validation Folds 10
Seed 1

(Num) quality

Start Stop

Result list (right-click for options)

11:33:11 - BestFirst + CfsSubsetEval
11:34:07 - Ranker + PrincipalComponent

Attribute selection output

```

0.1301 0.0563 -0.2755 -0.461 -0.6304 -0.0314 0.4017 0.3549 -0.0018 -0.0147 chlorides
-0.0739 0.6436 0.1471 -0.0317 0.0337 -0.0358 -0.0901 0.2404 0.2611 0.587 free sulfur dioxide
0.0014 0.6633 0.0006 0.0018 0.1254 -0.0246 -0.1346 0.1594 -0.2292 -0.5232 total sulfur dioxide
0.0433 0.0156 -0.1293 -0.4433 0.3103 0.8257 -0.0249 -0.0358 0.0619 -0.0011 density
-0.5259 -0.039 0.2262 -0.0912 -0.1155 0.1136 0.0793 -0.0866 -0.5893 0.3989 pH
0.2564 0.0679 0.3173 -0.4195 -0.3632 -0.0927 -0.5003 -0.5042 0.0577 0.0071 sulphates
-0.042 -0.2043 0.6848 0.1539 -0.2195 0.282 -0.0807 0.4786 0.2393 -0.1832 alcohol

```

Ranked attributes:

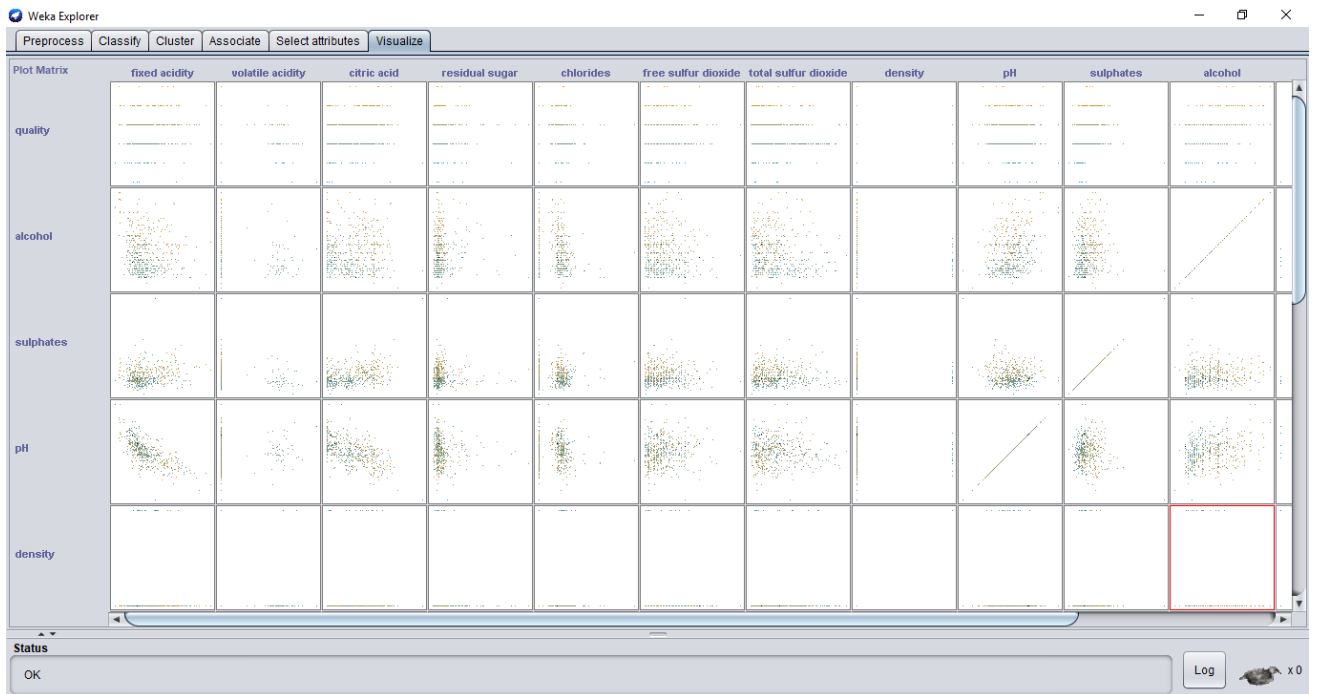
```

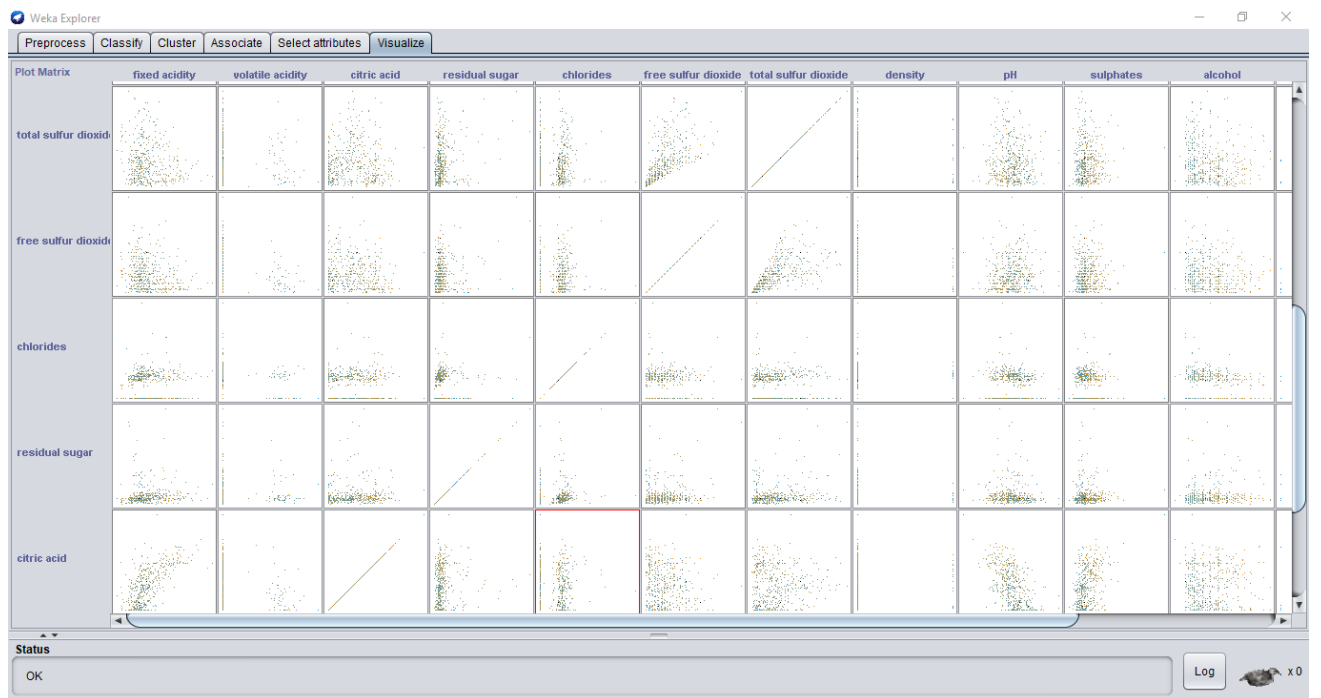
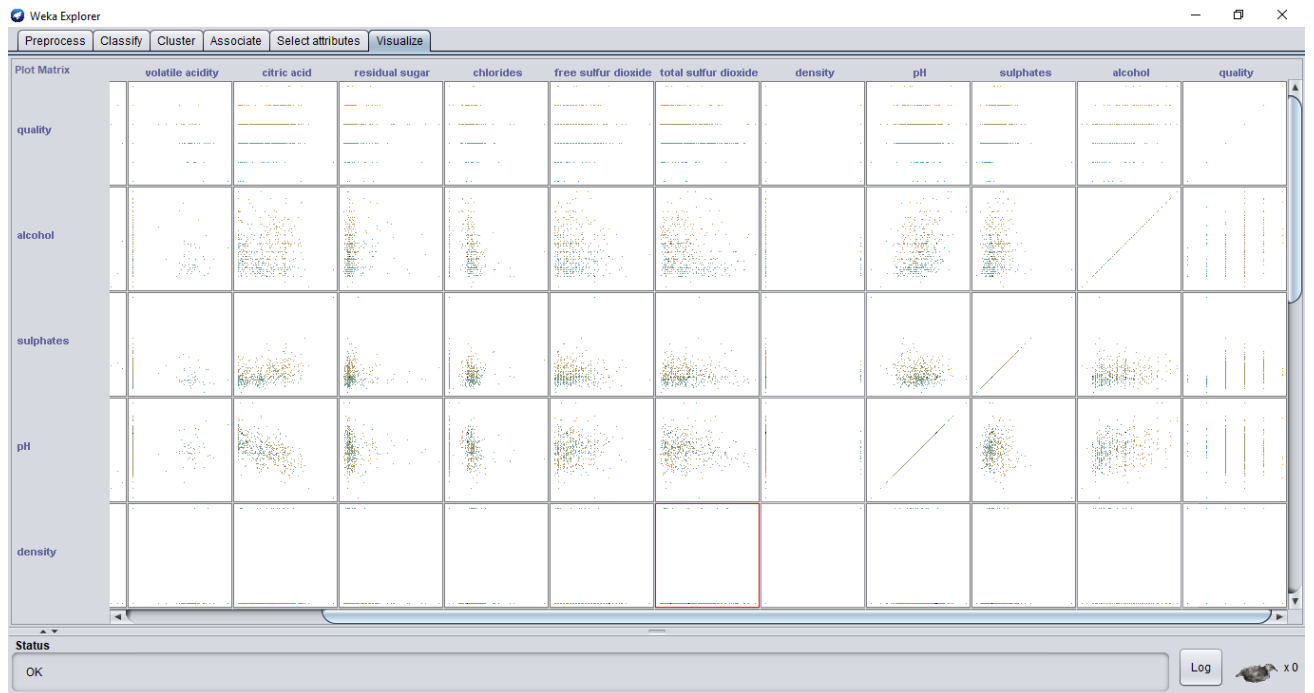
0.777 1 0.555fixed acidity+0.54 citric acid-0.526pH+0.256sulphates+0.13 chlorides...
0.6113 2 0.663total sulfur dioxide+0.644free sulfur dioxide+0.288residual sugar-0.204alcohol-0.098fixed acidity...
0.4978 3 0.685alcohol-0.462volatile acidity+0.317sulphates-0.276chlorides+0.226pH...
0.4008 4 0.489residual sugar-0.461chlorides-0.443density-0.42sulphates+0.329volatile acidity...
0.306 5 -0.63chlorides-0.382volatile acidity-0.365residual sugar-0.363sulphates+0.31 density...
0.2178 6 0.826density+0.385residual sugar+0.282alcohol+0.251volatile acidity+0.114pH...
0.1406 7 -0.625volatile acidity-0.9sulphates+0.402chlorides+0.396residual sugar-0.135total sulfur dioxide...
0.0848 8 -0.504sulphates+0.479alcohol-0.465residual sugar+0.355chlorides+0.24 free sulfur dioxide...
0.0507 9 -0.667citric acid-0.589pH+0.261free sulfur dioxide+0.239alcohol-0.229total sulfur dioxide...
0.0199 10 0.587free sulfur dioxide-0.523total sulfur dioxide+0.43 fixed acidity+0.399pH-0.183alcohol...

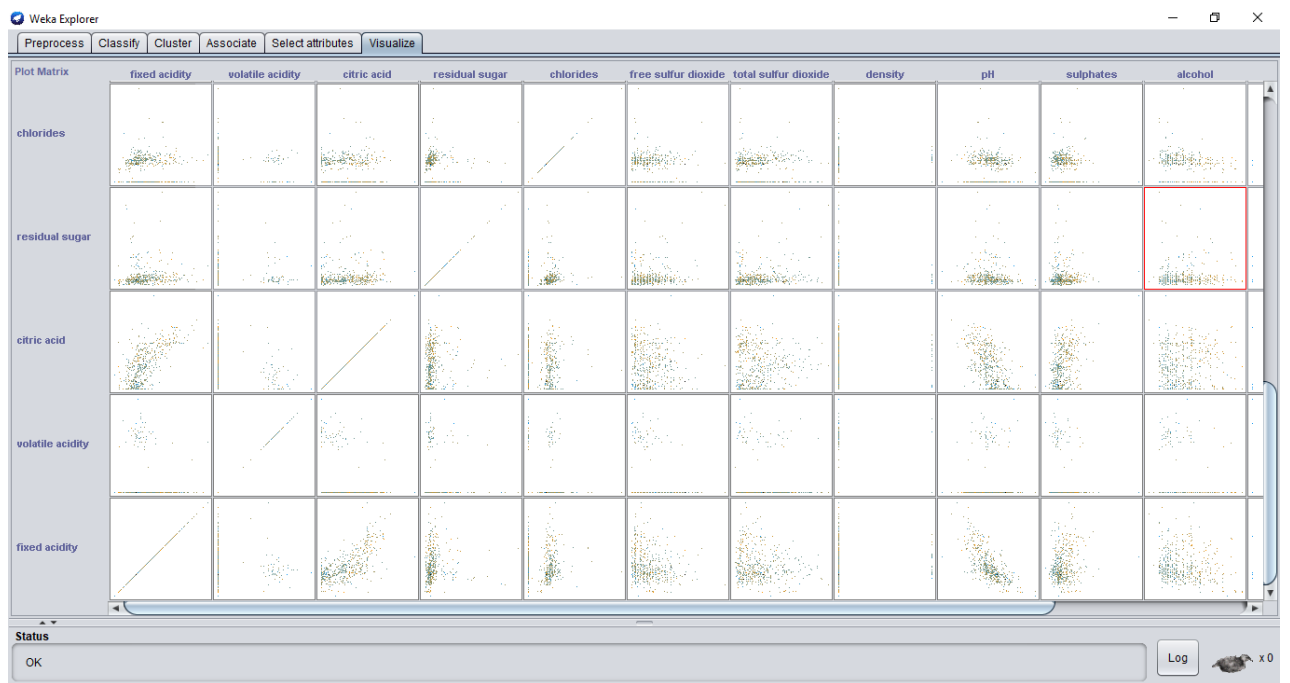
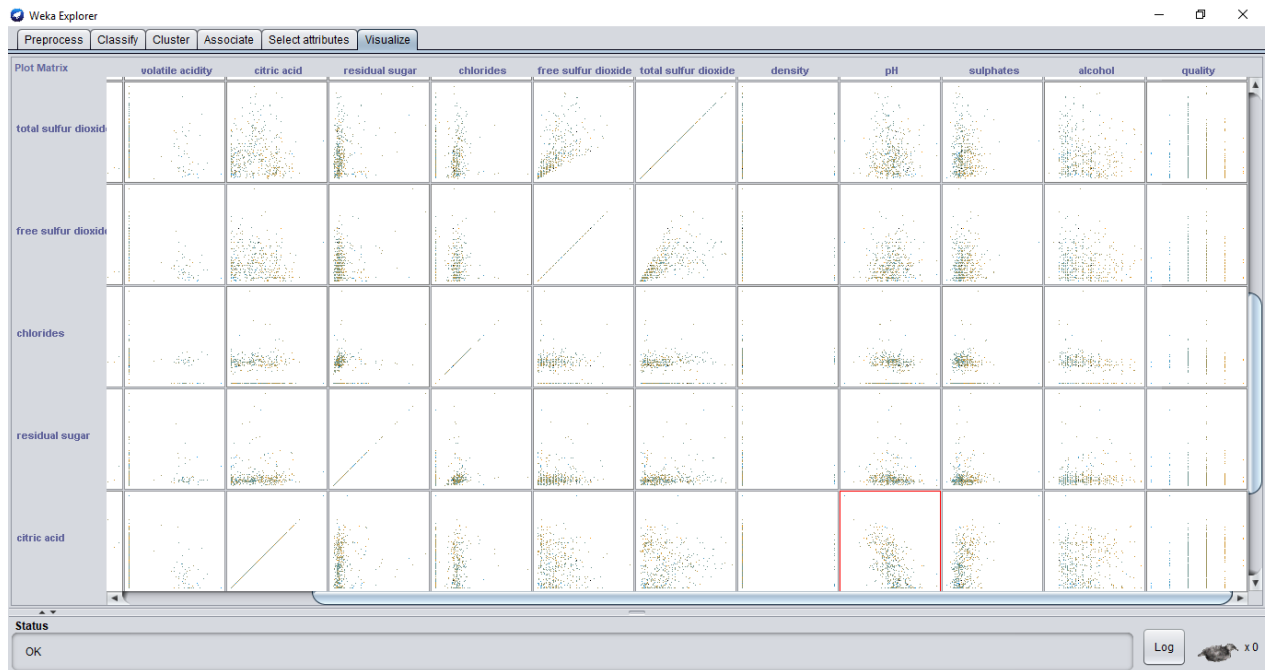
```

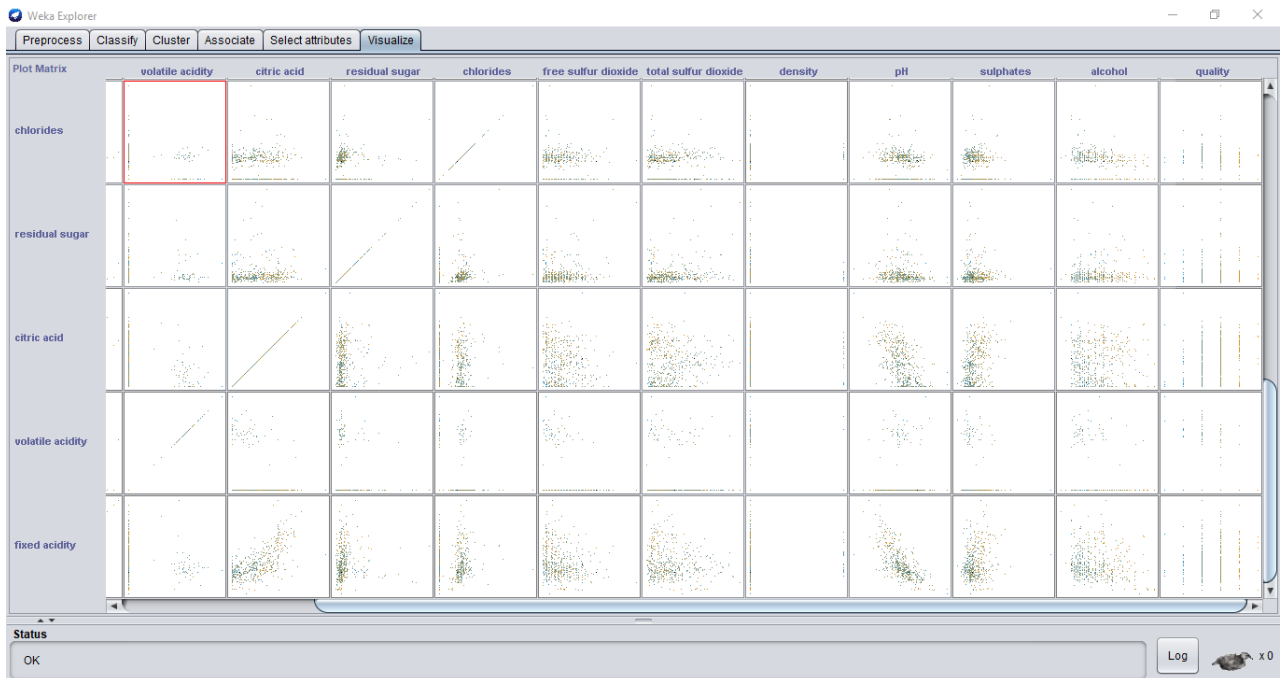
Selected attributes: 1,2,3,4,5,6,7,8,9,10 : 10

- Visualizar el scatter:

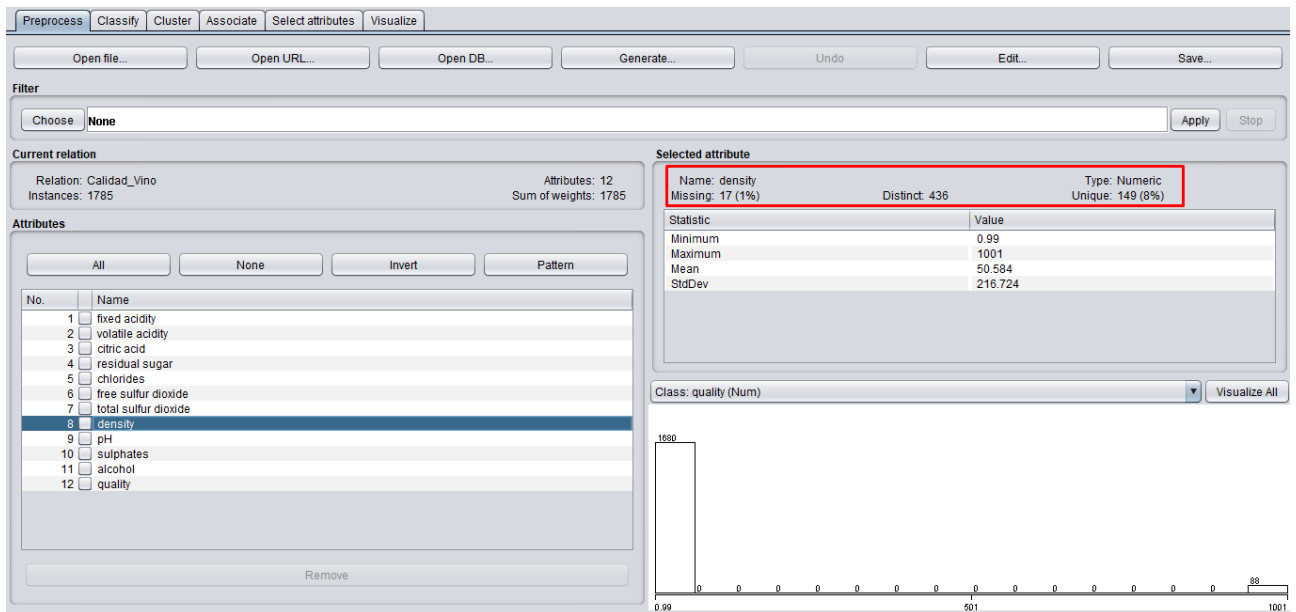


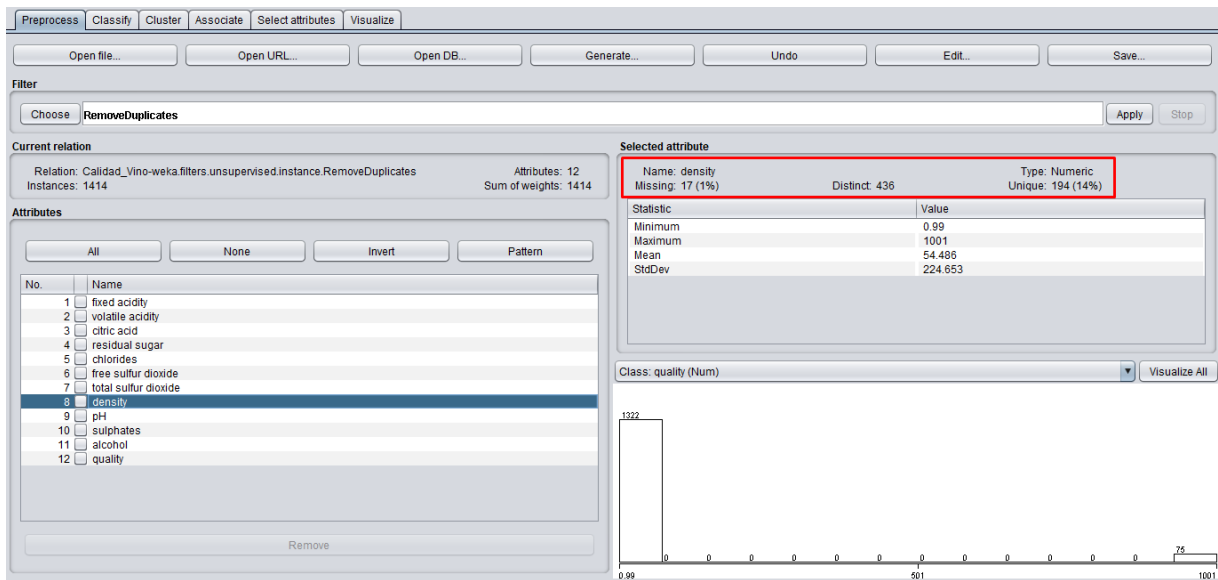




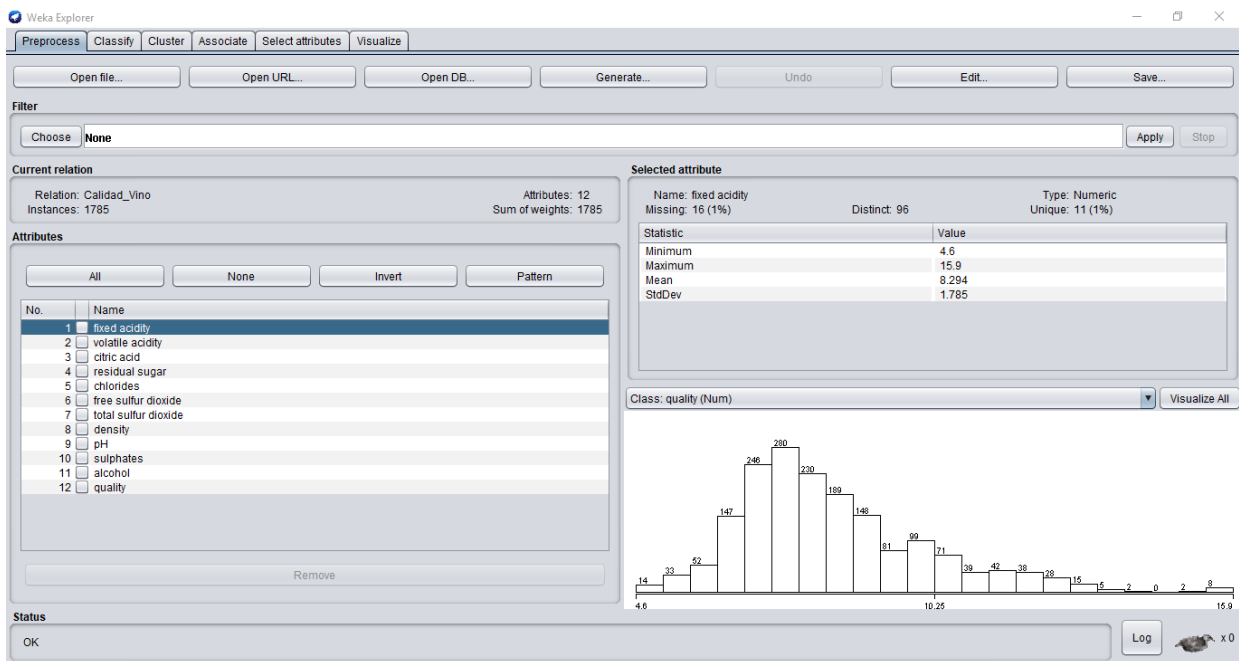


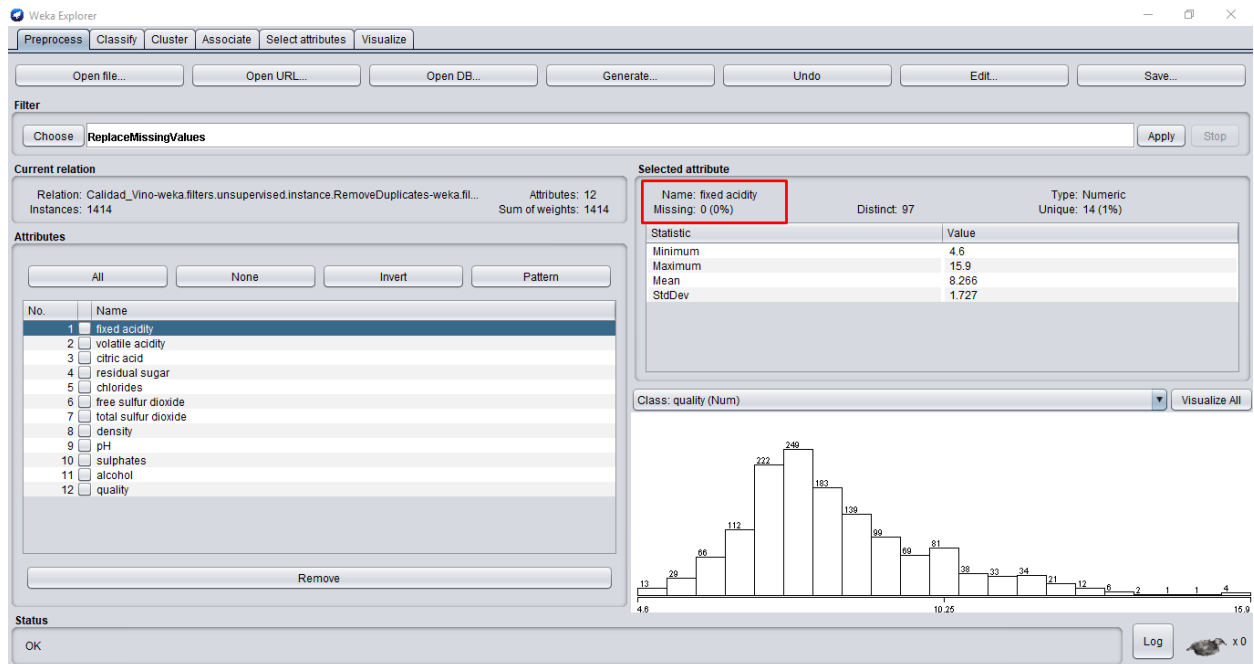
- Remover variables duplicadas:





- Remover valores faltantes:





- Visualizar de nuevo la matriz de correlación:

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (unsupervised):

Principal Components Attribute Transformer

Correlation matrix

1	-0.08	0.66	0.12	0.05	-0.15	-0.1	0.07	-0.68	0.19	-0.07
-0.08	1	-0.21	0.04	0.02	-0.07	-0.01	0	0.07	-0.08	-0.08
0.66	-0.21	1	0.15	0.1	-0.05	0.05	0.06	-0.55	0.32	0.1
0.12	0.04	0.15	1	0.03	0.15	0.2	-0.02	-0.09	-0.01	0.06
0.05	0.02	0.1	0.03	1	-0.01	-0.02	0.02	-0.14	0.18	-0.13
-0.15	-0.07	-0.05	0.15	-0.01	1	0.66	-0.02	0.06	0.05	-0.07
-0.1	-0.01	0.05	0.2	-0.02	0.66	1	0.03	-0.08	0.03	-0.21
0.07	0	0.06	-0.02	0.02	-0.02	0.03	1	-0.03	0.01	-0.04
-0.68	0.07	-0.55	-0.09	-0.14	0.06	-0.08	-0.03	1	-0.21	0.21
0.19	-0.08	0.32	-0.01	0.18	0.05	0.03	0.01	-0.21	1	0.09
-0.07	-0.08	0.1	0.06	-0.13	-0.07	-0.21	-0.04	0.21	0.09	1

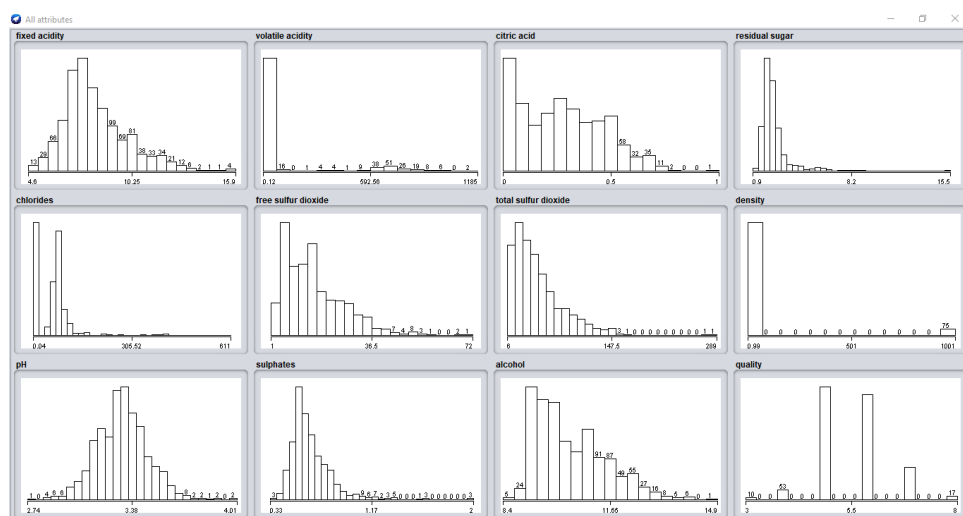
Ranked attributes:

0.7738	1	0.545	fixed acidity+0.538citric acid-0.524pH+0.28 sulphates+0.143chlorides...
0.6101	2	0.673	total sulfur dioxide+0.645free sulfur dioxide+0.263residual sugar-0.217alcohol-0.096fixed acidity...
0.4973	3	-0.712	alcohol+0.429volatile acidity+0.291chlorides-0.223sulphates-0.218pH...
0.3988	4	-0.638	chlorides-0.581sulphates+0.371residual sugar+0.243fixed acidity+0.172volatile acidity...
0.305	5	-0.563	residual sugar-0.55volatile acidity+0.415density-0.368chlorides-0.207alcohol...
0.2156	6	0.886	density+0.254residual sugar+0.248volatile acidity+0.241alcohol+0.142pH...
0.1392	7	0.588	volatile acidity-0.468chlorides+0.454sulphates-0.422residual sugar+0.125total sulfur dioxide...
0.0846	8	-0.545	sulphates+0.472alcohol-0.426residual sugar+0.352chlorides+0.241free sulfur dioxide...
0.0512	9	0.673	citric acid+0.549pH-0.304free sulfur dioxide+0.26 total sulfur dioxide-0.232alcohol...
0.0199	10	0.548	free sulfur dioxide-0.507total sulfur dioxide+0.454pH+0.428fixed acidity-0.213alcohol...

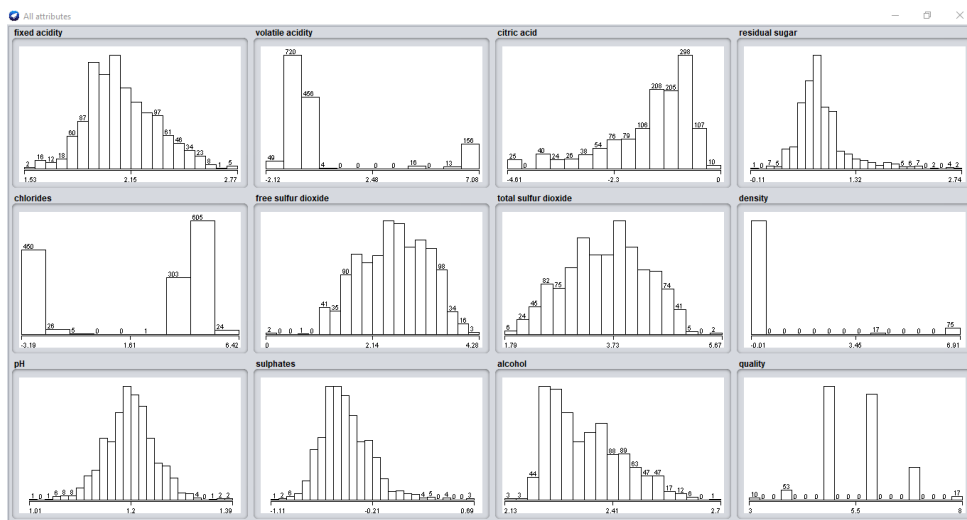
Selected attributes: 1,2,3,4,5,6,7,8,9,10 : 10

- Normalizar las variables para eliminar datos atípicos:

Antes:



Después:

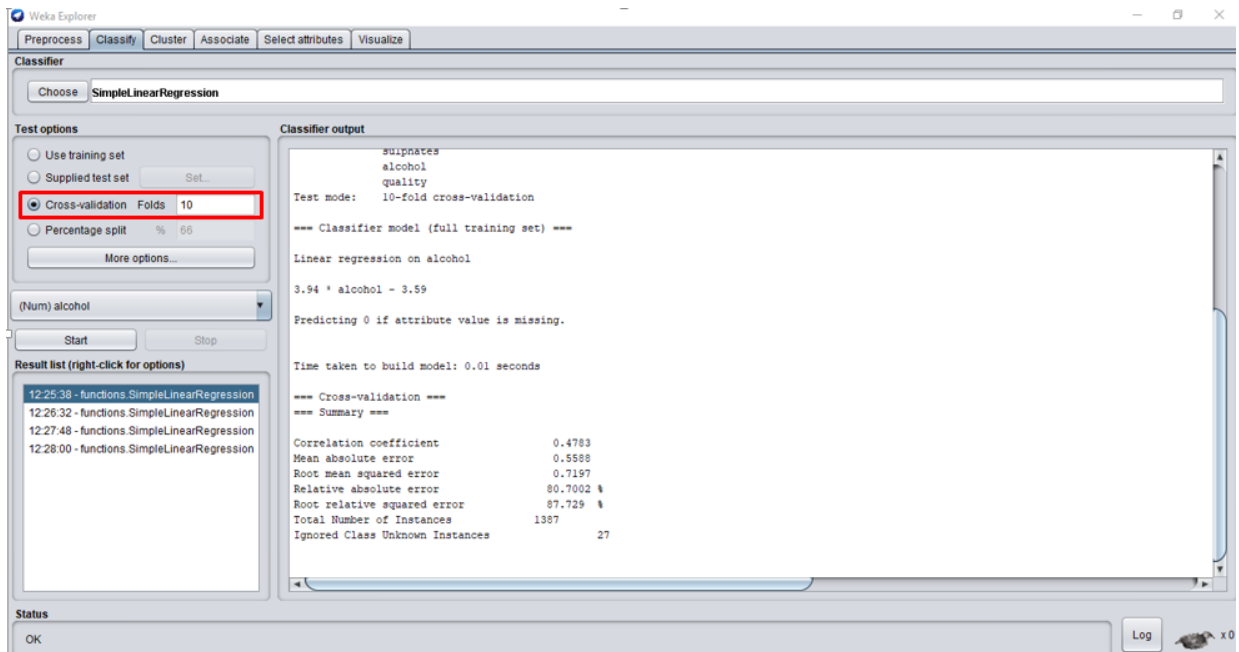


MÉTODOS SUPERVISADOS

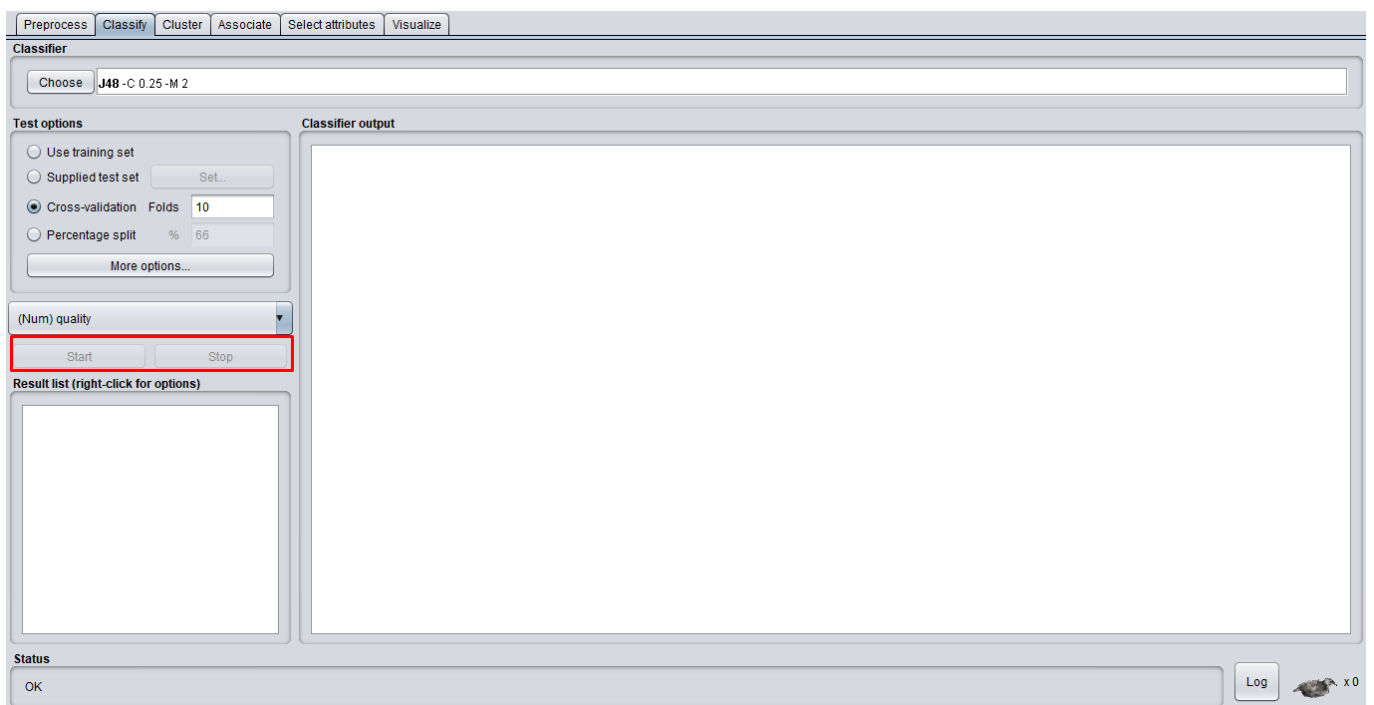
PREDICCIÓN:

A cada método se le aplica Cross-validation con Folds de 10.

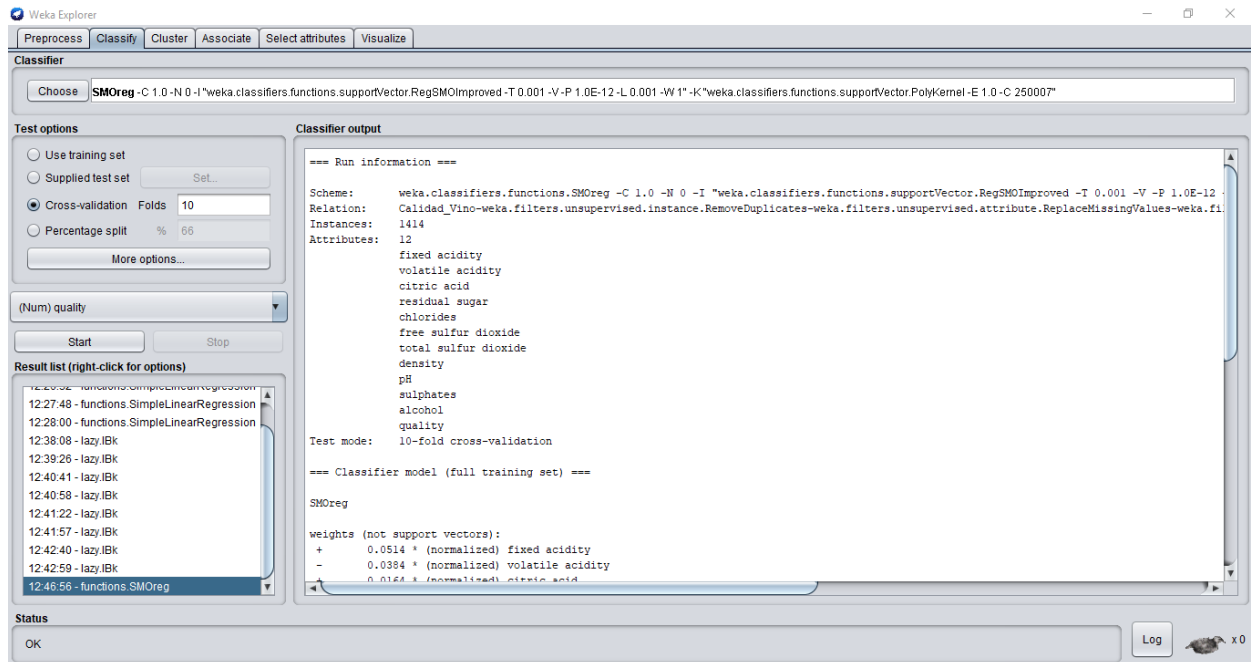
- REGRESIÓN LINEAL:



- DECISION TREE: El método no funciona



- SVM:



=== Classifier model (full training set) ===

SMOreg

weights (not support vectors):

```

+ 0.0514 * (normalized) fixed acidity
- 0.0384 * (normalized) volatile acidity
+ 0.0164 * (normalized) citric acid
+ 0.0019 * (normalized) residual sugar
- 0.0067 * (normalized) chlorides
+ 0.1144 * (normalized) free sulfur dioxide
- 0.1458 * (normalized) total sulfur dioxide
- 0.0082 * (normalized) density
- 0.0582 * (normalized) pH
+ 0.2813 * (normalized) sulphates
+ 0.4321 * (normalized) alcohol
+ 0.2614
  
```

Number of kernel evaluations: 1581649 (87.677% cached)

Time taken to build model: 0.89 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.5583
Mean absolute error	0.5133
Root mean squared error	0.6811
Relative absolute error	74.1252 %
Root relative squared error	83.0226 %
Total Number of Instances	1387
Ignored Class Unknown Instances	27

- RANDOM FOREST:

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 Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66

More options...

(Num) quality

Start Stop

Result list (right-click for options)

- 12:27:40 - functions.SimpleLinearRegression
- 12:38:08 - lazy.IBK
- 12:39:26 - lazy.IBK
- 12:40:41 - lazy.IBK
- 12:40:58 - lazy.IBK
- 12:41:22 - lazy.IBK
- 12:41:57 - lazy.IBK
- 12:42:40 - lazy.IBK
- 12:42:59 - lazy.IBK
- 12:46:56 - functions.SMOreg
- 12:52:43 - trees.RandomForest**

Classifier output

```

    pH
    sulphates
    alcohol
    quality
Test mode: 10-fold cross-validation

=== 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: 0.87 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.6293
Mean absolute error         0.4901
Root mean squared error     0.6371
Relative absolute error     70.7766 %
Root relative squared error 77.6553 %
Total Number of Instances   1387
Ignored Class Unknown Instances      27
  
```

Status

OK Log x0

- REDES NEURONALES:

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **MultilayerPerceptron** -L 0.5 -M 0.2 -N 1000 -V 0 -S 0 -E 20 -H a

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66

More options...

(Num) quality

Start Stop

Result list (right-click for options)

- 12:39:26 - lazy.IBK
- 12:40:41 - lazy.IBK
- 12:40:58 - lazy.IBK
- 12:41:22 - lazy.IBK
- 12:41:57 - lazy.IBK
- 12:42:40 - lazy.IBK
- 12:42:59 - lazy.IBK
- 12:46:56 - functions.SMOreg
- 12:52:43 - trees.RandomForest
- 12:57:17 - trees.RandomForest
- 13:05:41 - functions.MultilayerPerceptron**

Classifier output

```

=== Classifier model (full training set) ===

Linear Node 0
Inputs  Weights
Threshold  0.012695793562826851
Node 1    -3.2084646463027036
Node 2    -1.3165161933025595
Node 3    -1.4713866521144783
Node 4    -1.3050448492021722
Node 5    0.5447155267839326
Node 6    0.9895186615415692

Sigmoid Node 1
Inputs  Weights
Threshold  -8.761941327371606
Attrib fixed acidity  13.996850516779162
Attrib volatile acidity  1.1675221019815027
Attrib citric acid    -7.387590706380039
Attrib residual sugar  4.581913855783561
Attrib chlorides      4.569366955280723
Attrib free sulfur dioxide  -0.4050866490191467
Attrib total sulfur dioxide  -0.8145929208610925
Attrib density        4.773223944827658
Attrib pH             1.0886084522373989
Attrib sulphates      5.115613881372541
Attrib alcohol        2.585464405111567

Sigmoid Node 2
Inputs  Weights
  
```

Status

OK Log x0

Sigmoid Node 2

Inputs	Weights
Threshold	-4.4582187647328615
Attrib fixed acidity	-0.40563349185038833
Attrib volatile acidity	0.38152171453821054
Attrib citric acid	-0.14287853987183338
Attrib residual sugar	0.1203710803056117
Attrib chlorides	-0.24385525170380576
Attrib free sulfur dioxide	-0.11809116754141137
Attrib total sulfur dioxide	0.910065993319902
Attrib density	-0.07074849595974292
Attrib pH	-0.8825664082064211
Attrib sulphates	0.2405677066480379
Attrib alcohol	-3.917477268740148

Sigmoid Node 3

Inputs	Weights
Threshold	-15.622965477467762
Attrib fixed acidity	-5.152845927826507
Attrib volatile acidity	0.26351094575866046
Attrib citric acid	-0.702123649679872
Attrib residual sugar	5.831301562108013
Attrib chlorides	-5.863624141343358
Attrib free sulfur dioxide	-6.945431117627385
Attrib total sulfur dioxide	-3.86897149987818
Attrib density	-1.6314231057301567
Attrib pH	6.932647003888969
Attrib sulphates	-3.2629426967619573
Attrib alcohol	0.18653309320209305

Attrib free sulfur dioxide	-6.945431117627385
Attrib total sulfur dioxide	-3.86897149987818
Attrib density	-1.6314231057301567
Attrib pH	6.932647003888969
Attrib sulphates	-3.2629426967619573
Attrib alcohol	0.18653309320209305

Sigmoid Node 4

Inputs	Weights
Threshold	-6.019053323698511
Attrib fixed acidity	0.4139115077122772
Attrib volatile acidity	-1.148845115282567
Attrib citric acid	-0.4582379085529743
Attrib residual sugar	1.3132393643927647
Attrib chlorides	1.0304434576557744
Attrib free sulfur dioxide	0.3227842080923362
Attrib total sulfur dioxide	-1.5888857079315335
Attrib density	0.4062341582723168
Attrib pH	2.6126419690110674
Attrib sulphates	-6.3971314504976435
Attrib alcohol	0.023178605085421737

Sigmoid Node 5

Inputs	Weights
Threshold	-11.336210847742683
Attrib fixed acidity	0.1916720980259231
Attrib volatile acidity	-11.262973052452631
Attrib citric acid	-1.570416974169361
Attrib residual sugar	1.3917140686252036
Attrib chlorides	0.08611115191186484

```

Attrib chlorides      0.08611115191186484
Attrib free sulfur dioxide  2.400292572993844
Attrib total sulfur dioxide -2.6694629112133286
Attrib density        -2.02727021205916
Attrib pH             -1.3673652922028878
Attrib sulphates      0.8984799694911312
Attrib alcohol        3.0305405140055304
Sigmoid Node 6
Inputs  Weights
Threshold -2.819087750489133
Attrib fixed acidity  0.8405585530794177
Attrib volatile acidity  1.8302617803648014
Attrib citric acid    0.12029925528830675
Attrib residual sugar -3.2833847768113804
Attrib chlorides      -0.6122214332882024
Attrib free sulfur dioxide -4.697114610587977
Attrib total sulfur dioxide 0.4888348412169438
Attrib density        2.466668422813461
Attrib pH             5.862370602749059
Attrib sulphates      8.435860987812159
Attrib alcohol        5.814858175506903
Class
Input
Node 0

Time taken to build model: 2.64 seconds

```

```

Time taken to build model: 2.64 seconds

=== Cross-validation ===
=== Summary ===

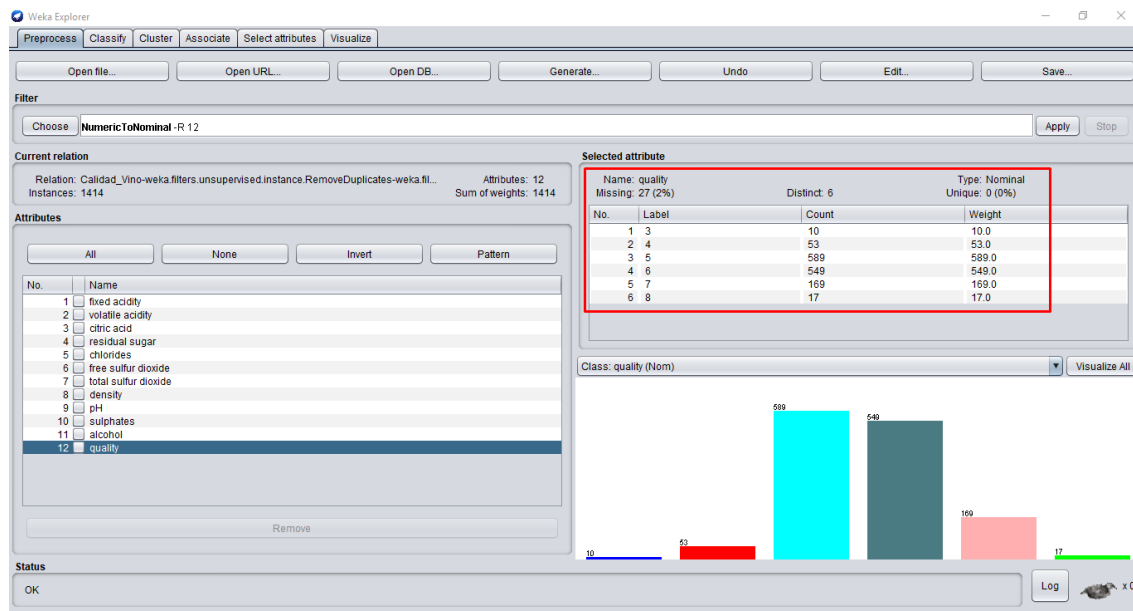
Correlation coefficient      0.3702
Mean absolute error         0.6503
Root mean squared error     0.8812
Relative absolute error     93.916 %
Root relative squared error 107.4137 %
Total Number of Instances   1387
Ignored Class Unknown Instances 27

```

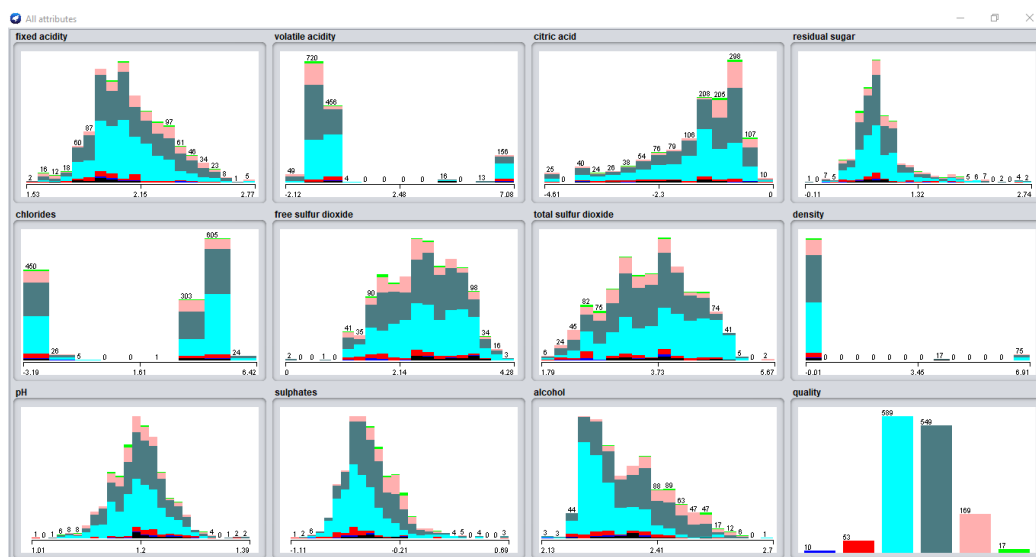
CLASIFICACIÓN:

- Convertir variables numéricas a categóricas:

Selected attribute		
Name: quality		Type: Numeric
Missing: 27 (2%)		Unique: 0 (0%)
Distinct: 6		
Statistic	Value	
Minimum	3	
Maximum	8	
Mean	5.624	
StdDev	0.82	



- Visualizar los datos:



- Balancear los datos:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **SMOTE -C 0 -K 5 -P 100.0 -S 1** Apply Stop

Current relation: Relation: Calidad_Vino-weka.filters.unsupervised.Instance Remove Duplicates-weka fil... Attributes: 12 Instances: 1502 Sum of weights: 1502

Attributes: All None Invert Pattern

No.	Name
1	fixed acidity
2	volatile acidity
3	citric acid
4	residual sugar
5	chlorides
6	free sulfur dioxide
7	total sulfur dioxide
8	density
9	pH
10	sulphates
11	alcohol
12	quality

Remove

Status: OK Log x 0

Selected attribute

Name: quality
Missing: 27 (2%)
Distinct: 6
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	3	47	47.0
2	4	53	53.0
3	5	589	589.0
4	6	549	549.0
5	7	169	169.0
6	8	69	68.0

Class: quality (Nom) Visualize All

- REGRESIÓN LOGÍSTICA:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier: Choose **Simple Logistic -I 0 -M 500 -H 50 -W 0.0**

Test options: Use training set, Supplied test set, **Cross-validation Folds 10**, Percentage split % 66, More options...

(Nom) quality Start Stop

Result list (right-click for options): 17:14:40 - functions.SimpleLogistic

Classifier output

Time taken to build model: 0.95 seconds

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

Correctly Classified Instances	802	54.3729 %
Incorrectly Classified Instances	673	45.6271 %
Kappa statistic	0.2834	
Mean absolute error	0.1918	
Root mean squared error	0.3109	
Relative absolute error	83.9884 %	
Root relative squared error	92.0417 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,000	0,000	?	0,000	?	?	0,780	0,124	3
	0,019	0,002	0,250	0,019	0,035	0,060	0,745	0,126	4
	0,740	0,291	0,628	0,740	0,680	0,441	0,793	0,678	5
	0,588	0,371	0,484	0,588	0,531	0,211	0,663	0,505	6
	0,166	0,038	0,364	0,166	0,228	0,184	0,831	0,322	7
	0,206	0,014	0,424	0,206	0,277	0,273	0,938	0,329	8
Weighted Avg.	0,544	0,260	?	0,544	?	?	0,754	0,519	

Status: OK Log x 0


```

=== Confusion Matrix ===

  a   b   c   d   e   f  <-- classified as
  0   2  24  21   0   0 |   a = 3
  0   1  28  23   0   1 |   b = 4
  0   1 436 147   3   2 |   c = 5
  0   0 196 323  23   7 |   d = 6
  0   0  10 122  28   9 |   e = 7
  0   0   0  31  23  14 |   f = 8

```

- NAIVE -BAYES:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) quality

Start Stop

Result list (right-click for options)

17:14:40 - functions SimpleLogistic

17:17:17 - bayes NaiveBayes

Classifier output

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	713	48.339 %
Incorrectly Classified Instances	762	51.661 %
Kappa statistic	0.2652	
Mean absolute error	0.1971	
Root mean squared error	0.3368	
Relative absolute error	86.3133 %	
Root relative squared error	99.7312 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,043	0,029	0,047	0,043	0,044	0,014	0,712	0,073	3
	0,113	0,027	0,133	0,113	0,122	0,093	0,720	0,089	4
	0,667	0,258	0,632	0,667	0,649	0,405	0,780	0,672	5
	0,415	0,238	0,509	0,415	0,457	0,187	0,626	0,474	6
	0,160	0,038	0,351	0,160	0,220	0,174	0,741	0,246	7
	0,938	0,130	0,238	0,938	0,370	0,402	0,915	0,341	8
Weighted Avg.	0,483	0,204	0,499	0,483	0,477	0,274	0,720	0,494	

Status

OK Log x0

```

=== Confusion Matrix ===

  a   b   c   d   e   f  <-- classified as
  2   6  24  14   0   1 |   a = 3
  2   6  27  15   3   0 |   b = 4
 18  18 393 135  11  14 |   c = 5
 19  11 167 228  33  91 |   d = 6
  2   4  10  49  27  77 |   e = 7
  0   0   1   7   3  57 |   f = 8

```

- SVM:

The screenshot shows the Weka Explorer interface with the Classifier tab selected. The classifier chosen is SMO (Sequential Minimal Optimization) with the following parameters: `-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 -R 1.0E-8 -M -1 -num-decimal-places 4"`.

Test options:

- Use training set: ☐
- Supplied test set: ☐ Set...
- Cross-validation: ☒ Folds: 10
- Percentage split: ☐ % 66

Classifier output:

Time taken to build model: 0.29 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	786	53.2881 %
Incorrectly Classified Instances	689	46.7119 %
Kappa statistic	0.2397	
Mean absolute error	0.2383	
Root mean squared error	0.3352	
Relative absolute error	104.3521 %	
Root relative squared error	99.242 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,000	0,000	?	0,000	?	?	?	0,712	0,056	3
0,000	0,000	?	0,000	?	?	?	0,639	0,050	4
0,761	0,316	0,615	0,761	0,680	0,436	?	0,752	0,577	5
0,616	0,442	0,452	0,616	0,522	0,168	?	0,585	0,413	6
0,000	0,000	?	0,000	?	?	?	0,803	0,273	7
0,000	0,000	?	0,000	?	?	?	0,909	0,226	8
Weighted Avg.	0,533	0,291	?	0,533	?	?	0,697	0,429	

Result list (right-click for options):

- 17:14:40 - functions.SimpleLogistic
- 17:17:17 - bayes.NaiveBayes
- 17:19:49 - functions.SMO

Status: OK

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
0	0	28	19	0	0	a = 3
0	0	29	24	0	0	b = 4
0	0	448	141	0	0	c = 5
0	0	211	338	0	0	d = 6
0	0	12	157	0	0	e = 7
0	0	0	68	0	0	f = 8

- DECISION TREE:

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
9	1	17	19	1	0	a = 3
1	2	30	20	0	0	b = 4
16	16	372	166	18	1	c = 5
13	18	182	262	65	9	d = 6
1	4	25	78	51	10	e = 7
0	0	1	12	16	39	f = 8

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
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

(Nom) quality

Result list (right-click for options)

- 17:14:40 - functions.SimpleLogistic
- 17:17:17 - bayes.NaiveBayes
- 17:19:49 - functions.SMO
- 17:23:13 - trees.J48**

Classifier output

Time taken to build model: 0.22 seconds

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

Correctly Classified Instances	735	49.8305 %
Incorrectly Classified Instances	740	50.1695 %
Kappa statistic	0.2559	
Mean absolute error	0.1763	
Root mean squared error	0.377	
Relative absolute error	77.2042 %	
Root relative squared error	111.6228 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,191	0,022	0,225	0,191	0,207	0,184	0,609	0,127	3
	0,038	0,027	0,049	0,038	0,043	0,012	0,555	0,045	4
	0,632	0,288	0,593	0,632	0,612	0,341	0,687	0,528	5
	0,477	0,319	0,470	0,477	0,474	0,158	0,566	0,419	6
	0,302	0,077	0,338	0,302	0,319	0,237	0,703	0,256	7
	0,574	0,014	0,661	0,574	0,614	0,599	0,825	0,449	8
Weighted Avg.	0,498	0,245	0,490	0,498	0,494	0,256	0,643	0,423	

Status

OK x0

- RANDOM FOREST:

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 **10**
☐ Percentage split % **66**

(Nom) quality

Result list (right-click for options)

- 17:14:40 - functions.SimpleLogistic
- 17:17:17 - bayes.NaiveBayes
- 17:19:49 - functions.SMO
- 17:23:13 - trees.J48
- 17:24:12 - 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: 1.28 seconds

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

Correctly Classified Instances	875	59.322 %
Incorrectly Classified Instances	600	40.678 %
Kappa statistic	0.3741	
Mean absolute error	0.1779	
Root mean squared error	0.2977	
Relative absolute error	77.9131 %	
Root relative squared error	88.1431 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,021	0,003	0,200	0,021	0,038	0,056	0,910	0,269	3
	0,000	0,001	0,000	0,000	0,000	-0,005	0,747	0,107	4
	0,747	0,277	0,642	0,747	0,691	0,462	0,816	0,725	5
	0,601	0,303	0,540	0,601	0,569	0,292	0,702	0,522	6
	0,308	0,041	0,491	0,308	0,378	0,328	0,859	0,457	7
	0,765	0,011	0,776	0,765	0,770	0,759	0,990	0,852	8
Weighted Avg.	0,593	0,229	0,556	0,593	0,568	0,367	0,787	0,588	

Status

OK x0

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
1	0	31	15	0	0	0	a = 3
0	0	33	19	1	0	0	b = 4
3	0	440	142	4	0	0	c = 5
1	1	173	330	39	5	0	d = 6
0	0	8	99	52	10	0	e = 7
0	0	0	6	10	52	0	f = 8

- REDES NEURONALES:

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
0	1	32	14	0	0	0	a = 3
0	0	36	16	1	0	0	b = 4
4	3	420	156	6	0	0	c = 5
0	6	188	300	43	12	0	d = 6
0	4	13	95	39	18	0	e = 7
0	0	1	19	19	29	0	f = 8

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: MultilayerPerceptron - L 0.3 - M 0.2 - N 500 - V 0 - S 0 - E 20 - H a

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) quality

Start Stop

Result list (right-click for options)

17:53:21 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 2.52 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	788	53.4237 %
Incorrectly Classified Instances	687	46.5763 %
Kappa statistic	0.2847	
Mean absolute error	0.1817	
Root mean squared error	0.3216	
Relative absolute error	79.5834 %	
Root relative squared error	55.2057 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

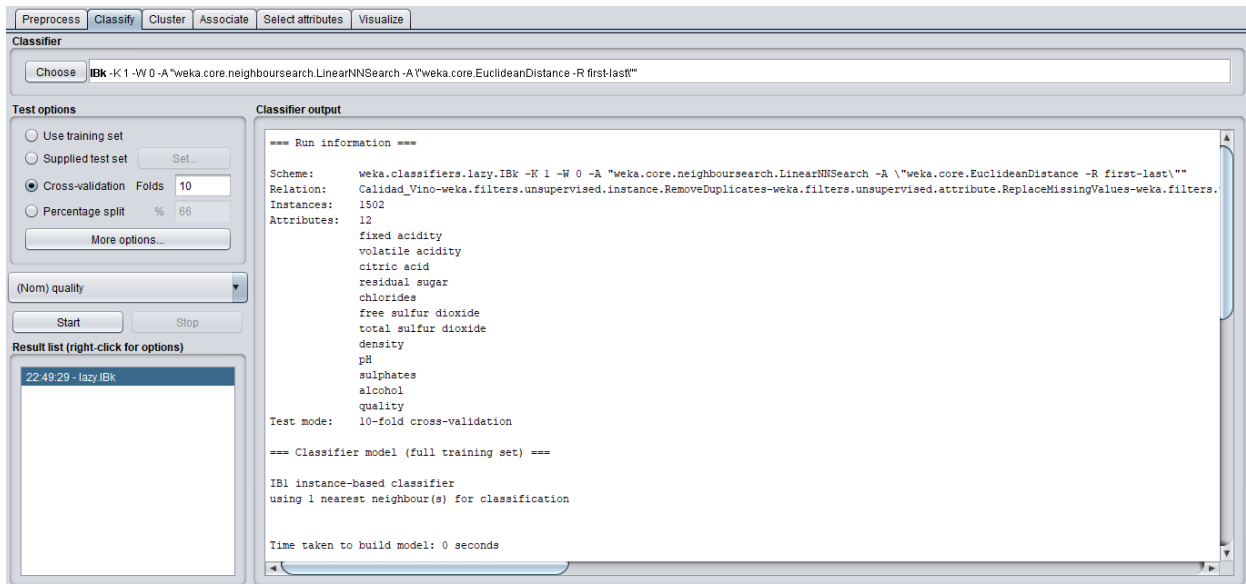
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,000	0,003	0,000	0,000	0,000	0,000	-0,009	0,729	0,071	3
0,000	0,010	0,000	0,000	0,000	0,000	-0,019	0,672	0,074	4
0,713	0,305	0,609	0,713	0,657	0,401	0,771	0,654	0,454	5
0,546	0,324	0,500	0,546	0,522	0,219	0,654	0,482	0,482	6
0,231	0,053	0,361	0,231	0,282	0,218	0,786	0,319	0,319	7
0,426	0,021	0,492	0,426	0,457	0,434	0,871	0,397	0,397	8
Weighted Avg.	0,534	0,250	0,493	0,534	0,510	0,285	0,729	0,501	

Status

OK Log x 0

- KNN:



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

Correctly Classified Instances	745	50.5085 %
Incorrectly Classified Instances	730	49.4915 %
Kappa statistic	0.2816	
Mean absolute error	0.1655	
Root mean squared error	0.4053	
Relative absolute error	72.4599 %	
Root relative squared error	119.9862 %	
Total Number of Instances	1475	
Ignored Class Unknown Instances	27	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,383	0,022	0,367	0,383	0,375	0,354	0,652	0,139	3
	0,075	0,023	0,111	0,075	0,090	0,064	0,541	0,045	4
	0,560	0,269	0,581	0,560	0,570	0,294	0,647	0,509	5
	0,468	0,313	0,470	0,468	0,469	0,155	0,574	0,414	6
	0,444	0,087	0,397	0,444	0,419	0,340	0,674	0,241	7
	0,897	0,018	0,709	0,897	0,792	0,787	0,942	0,624	8
Weighted Avg.	0,505	0,236	0,501	0,505	0,502	0,264	0,633	0,420	

=== Confusion Matrix ===

	a	b	c	d	e	f	<-- classified as
18	0	17	12	0	0	0	a = 3
3	4	25	18	3	0	0	b = 4
15	20	330	189	34	1	1	c = 5
13	11	179	257	74	15	1	d = 6
0	1	16	68	75	9	1	e = 7
0	0	1	3	3	61	1	f = 8

CONCLUSIÓN:

- Predicción:

METODO	MEAN ABSOLUTE ERROR	ROOT MEAN SQUARED ERROR
Regresión Lineal	0.5588	0.7197
SVM	0.5133	0.6811
Random Forest	0.4901	0.6371
Redes Neuronales	0.6503	0.8812

Para los métodos de predicción aplicados el que presenta un mayor Root mean squared error y mean absolute error son las redes neuronales y el menor es Random Forest. Se puede concluir que el mejor método usado para la predicción de la calidad de vino es el Random Forest, ya que presenta valores de error menores a los demás.

- Clasificación:

METODO	MEAN ABSOLUTE ERROR	ROOT MEAN SQUARED ERROR
Regresión Logística	0.1971	0.3368
Naive-Bayes	0.1971	0.3368
SVM	0.2383	0.3352
Decision tree	0.1763	0.377
Random Forest	0.1779	0.2977
Redes Neuronales	0.1817	0.3216
KNN	0.1655	0.4653

Para los métodos de clasificación aplicados el que presenta un mayor mean absolute error es SVM y con el mayor Root mean squared error es KNN; el menor mean absolute error es KNN y el menor Root mean squared error es Random Forest. Se puede concluir que el mejor método usado para la predicción de la calidad de vino es el Random Forest, ya que presenta valores de error menores a los demás.