

SISTEMAS DE LA INFORMACIÓN

MINERÍA DE DATOS CON WEKA

Diego Santolaya Martínez

Abel Chils Trabanco

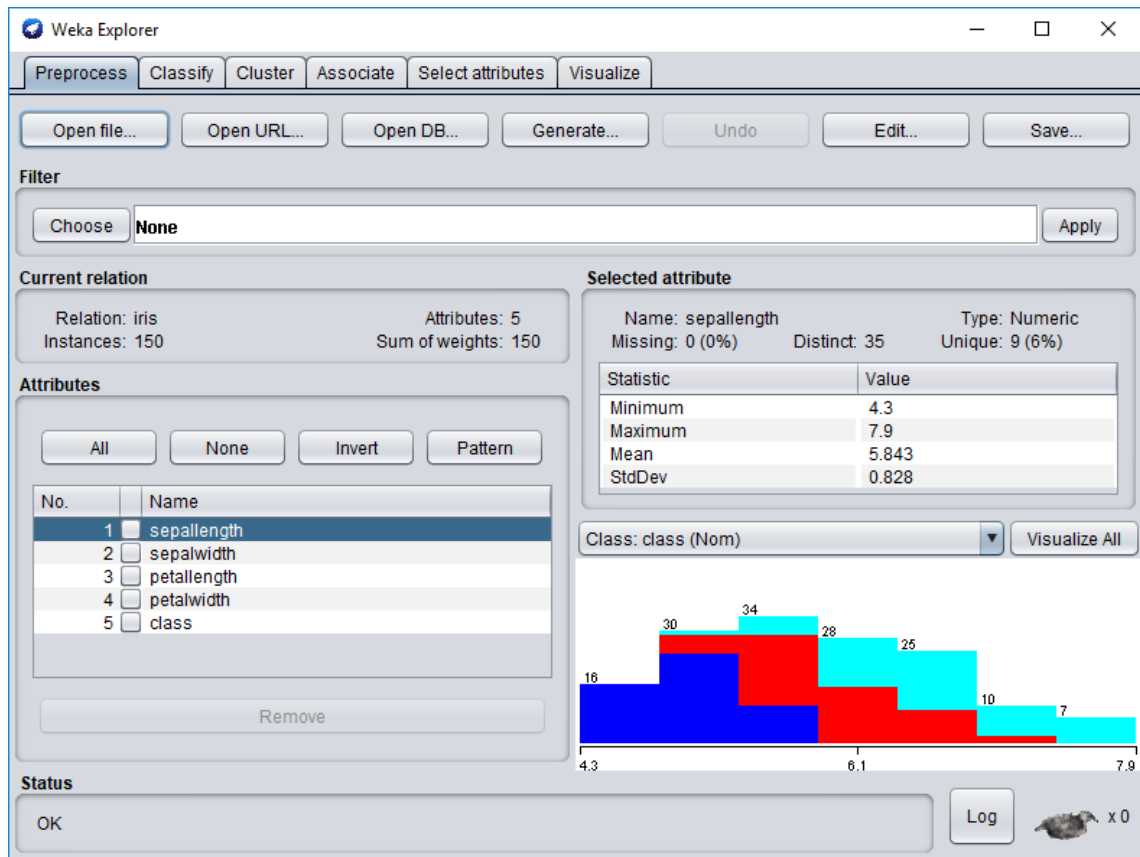
Alexandru Oarga Hategan

ÍNDICE:

1.
Introducción.....
.....pg 1
2. Primeros Pasos.....
.....pg 1
3. Búsqueda basada en palabras
clave....pg 2
4. Conclusiones.....
.....pg 3

1. Introducción

2. Primeros Pasos



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

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

(Nom) class

Start Stop

Result list (right-click for options)

12:00:08 - rules.ZeroR

Classifier output

```

Root relative squared error      100 %
Total Number of Instances      150

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC
1,000    1,000    0,333    1,000    0,500    0,000
0,000    0,000    0,000    0,000    0,000    0,000
0,000    0,000    0,000    0,000    0,000    0,000
Weighted Avg.    0,333    0,333    0,111    0,333    0,167    0,000

=== Confusion Matrix ===

  a  b  c  <-- classified as
50  0  0 | a = Iris-setosa
50  0  0 | b = Iris-versicolor
50  0  0 | c = Iris-virginica
  
```

Status

OK Log x0

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 % 66
 More options...

(Nom) class

Start Stop

Result list (right-click for options)

12:00:08 - rules.ZeroR
12:07:03 - trees.J48

Classifier output

```

Root relative squared error      33.6353 %
Total Number of Instances      150

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC
0,980    0,000    1,000    0,980    0,990    0,985
0,940    0,030    0,940    0,940    0,940    0,910
0,960    0,030    0,941    0,960    0,950    0,925
Weighted Avg.    0,960    0,020    0,960    0,960    0,960    0,940

=== Confusion Matrix ===

  a  b  c  <-- classified as
49  1  0 | a = Iris-setosa
0 47  3 | b = Iris-versicolor
0  2 48 | c = Iris-virginica
  
```

Status

OK Log x0

Weka Explorer

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

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

Filter

Choose **None** Apply

Current relation

Relation: weather.symbolic
Instances: 14

Attributes: 5
Sum of weights: 14

Attributes

All None Invert Pattern

Selects all attributes that match a reg. expression

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> windy
5	<input type="checkbox"/> play

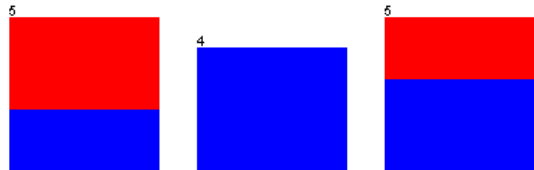
Remove

Selected attribute

Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) Visualize All



Status

OK Log x 0

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 % 66
More options...

(Nom) play

Start Stop

Result list (right-click for options)

- 12:00:08 - rules.ZeroR
- 12:07:03 - trees.J48
- 12:13:02 - trees.J48**

Classifier output

```

Root mean squared error          0
Relative absolute error          0 %
Root relative squared error      0 %
Total Number of Instances       14

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC
              1,000    0,000    1,000    1,000    1,000    1,000
              1,000    0,000    1,000    1,000    1,000    1,000
Weighted Avg.  1,000    0,000    1,000    1,000    1,000    1,000

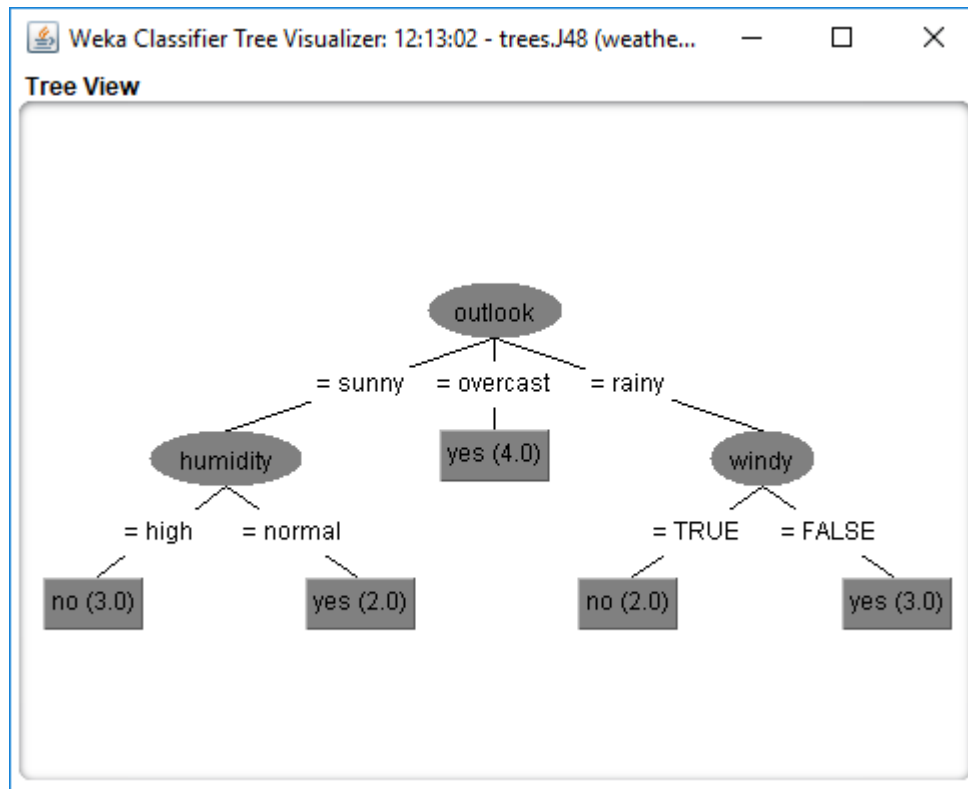
=== Confusion Matrix ===

 a b  <-- classified as
 9 0 | a = yes
 0 5 | b = no

```

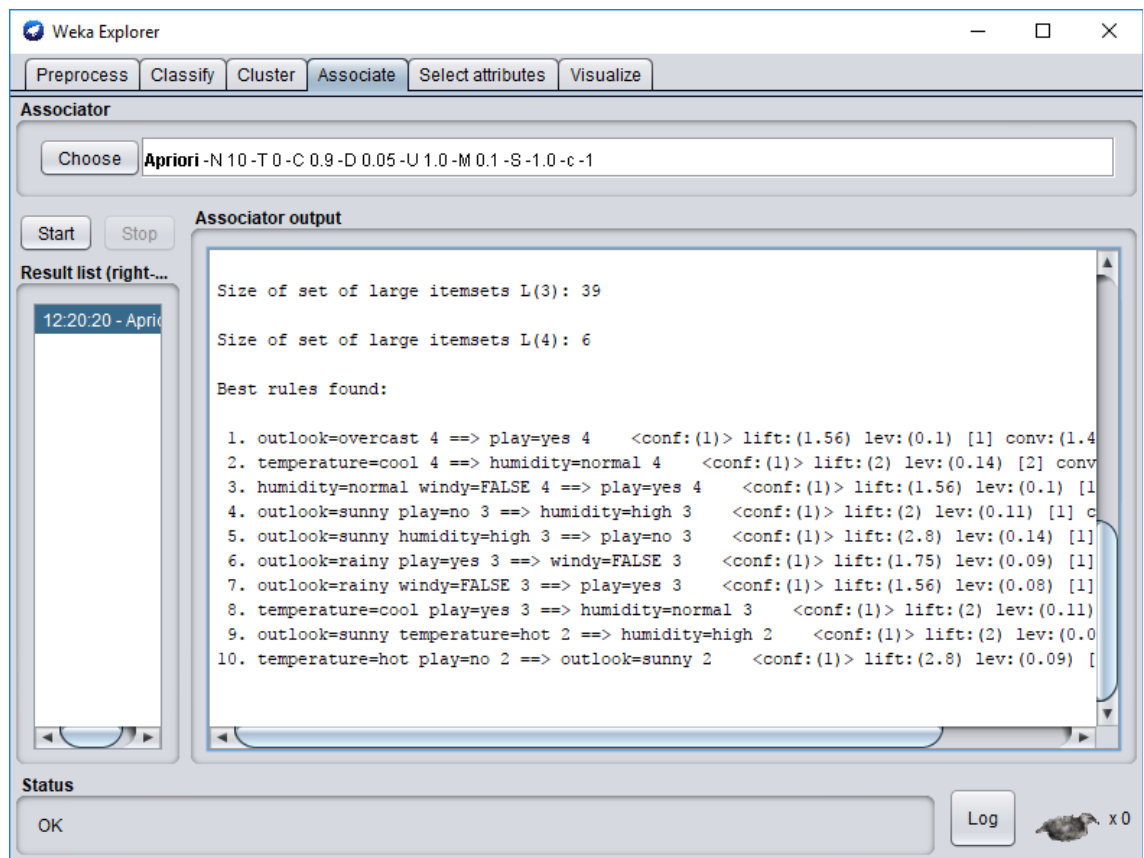
Status

OK Log x 0



Conclusiones:

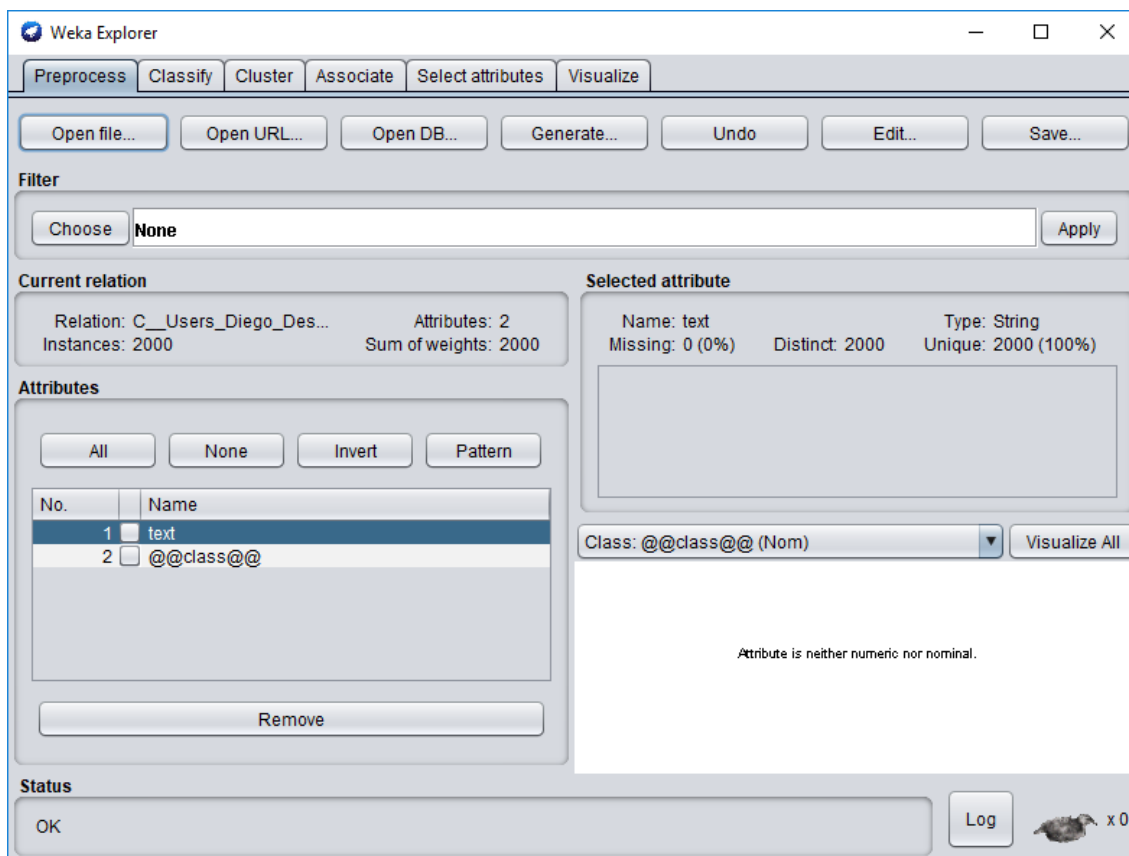
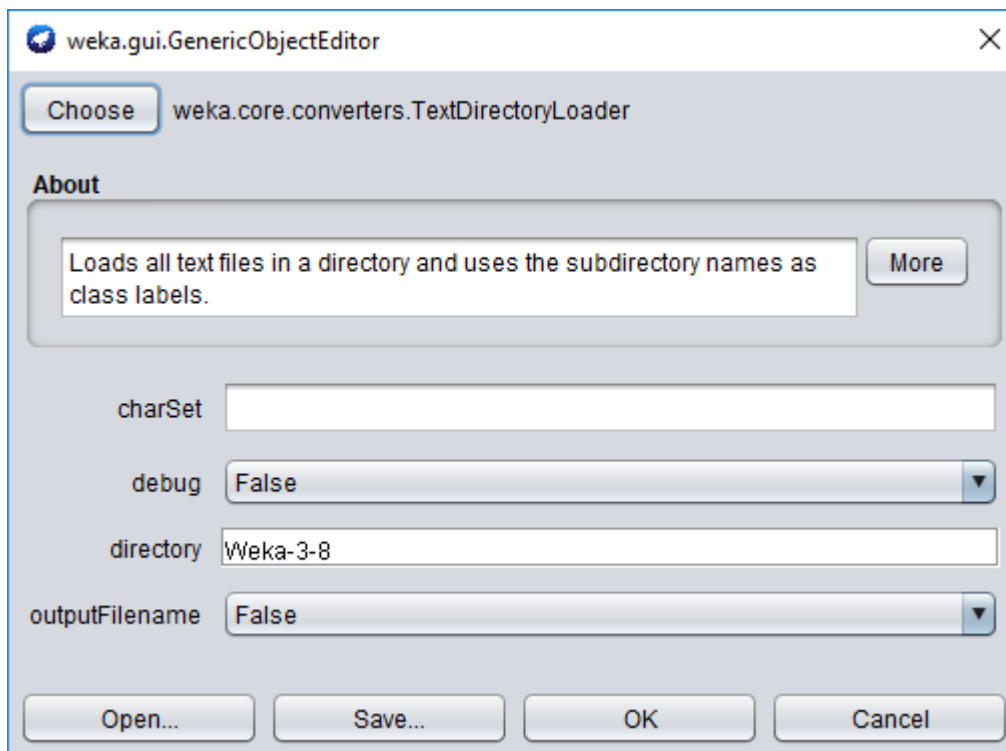
Los días que hace humedad alta o viento no se juega el partido, mientras que si está nublado o hace humedad normal sí que se disputan los encuentros.

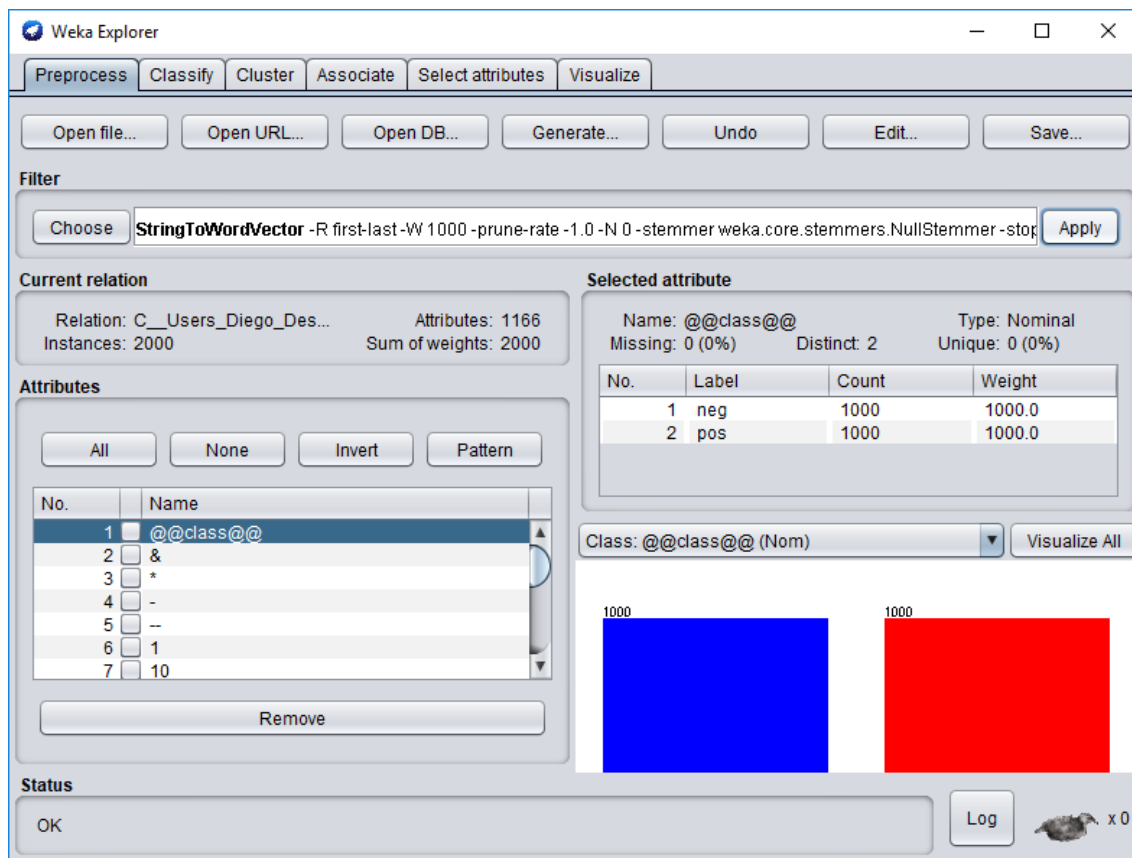


Conclusiones:

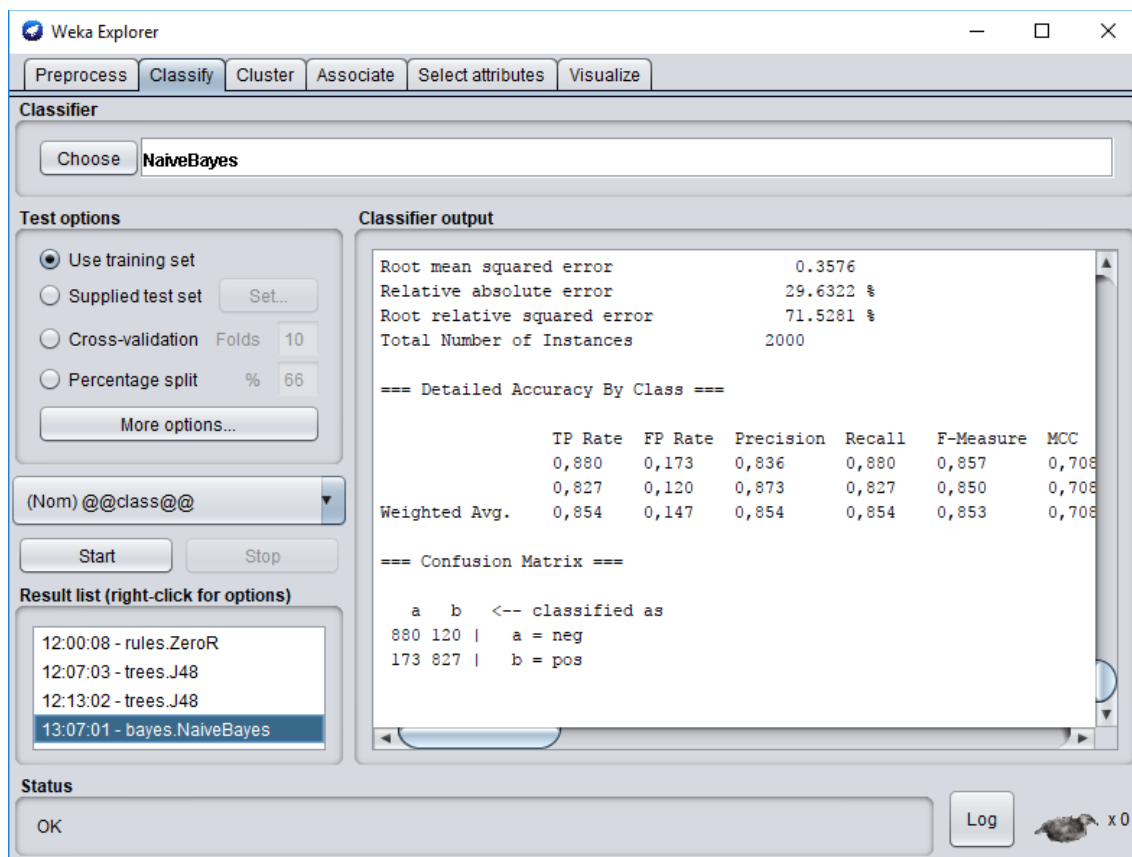
Las reglas obtenidas encajan con los resultados del algoritmo anterior, ya que tiempo nublado implicaba que siempre se jugaba el partido, y así lo refleja la primera regla. El resto de reglas son relaciones entre los distintos parámetros, como se explican a continuación: cuando hace frío la humedad no es alta, cuando no hay humedad ni viento se suelen jugar los partidos, si hace soleado y se ha suspendido el partido hay humedad, si hay lluvia y se juega el partido no hace viento, si hace frío y se juega la humedad es normal, etc. Estas reglas se hayan cuando hay varios casos en los que se repiten patrones teniendo los mismos parámetros.

8.

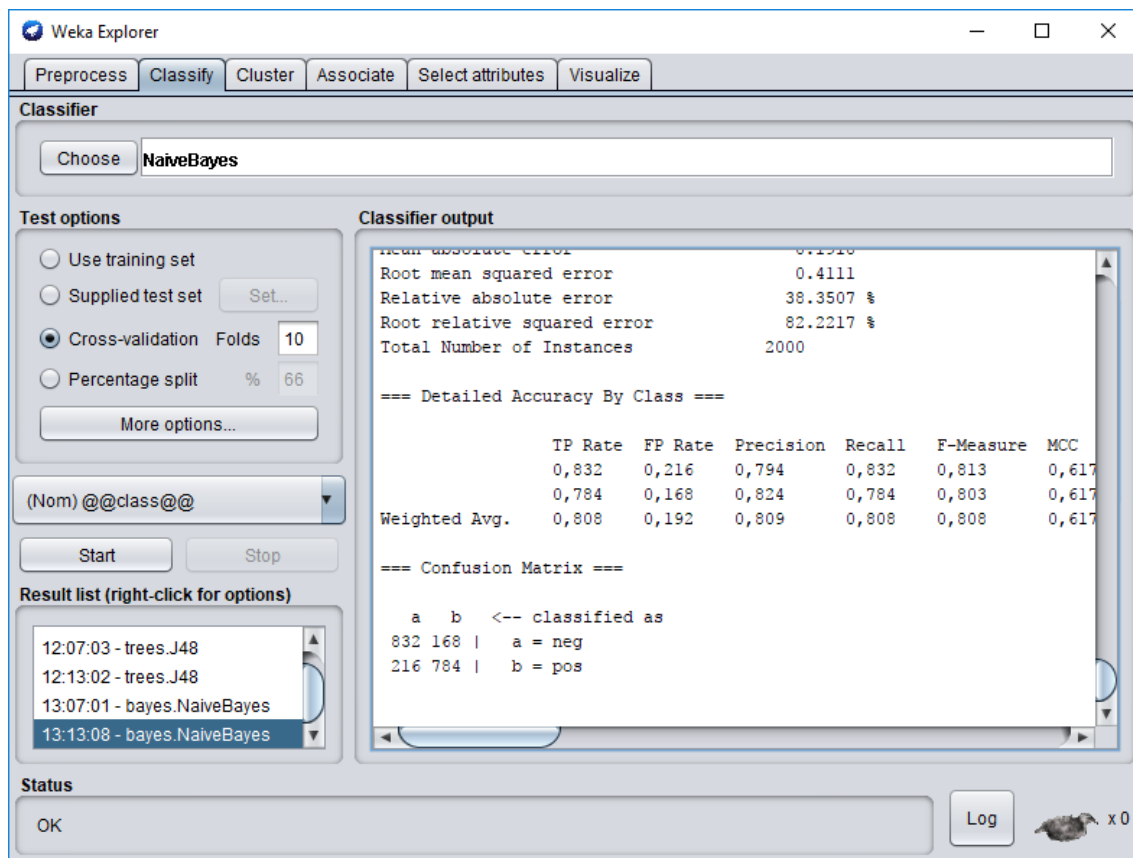




Use training set:

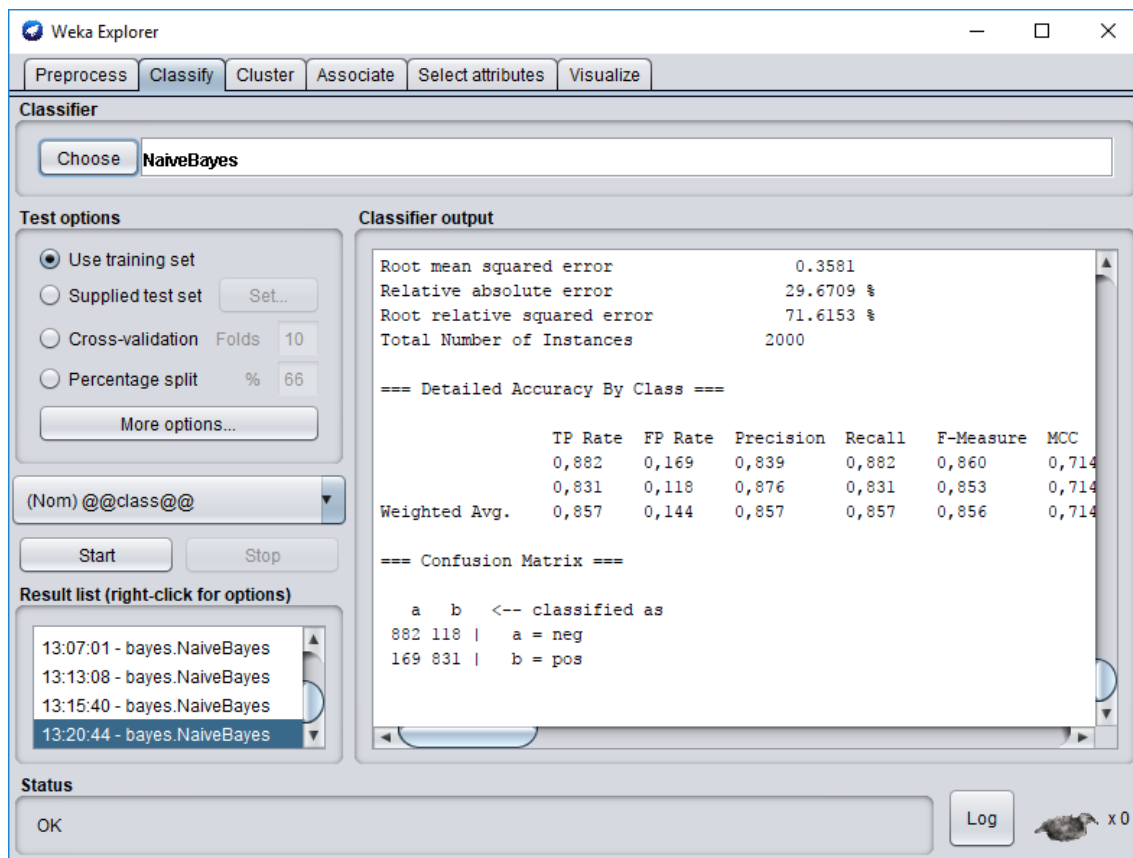


Cross-validation



- i. Precisión al detectar las opiniones positivas 87,3%
- ii. Precisión al detectar las opiniones negativas. 83,6%
- iii. Recall al detectar las opiniones positivas. 82,7%
- iv. Recall al detectar las opiniones negativas. 88,0%
- v. Precisión promedio. 85,4%
- vi. Recall promedio. 85,4%

tras eliminar palabras sobrantes:



- i. Precisión al detectar las opiniones positivas 87,6%
- ii. Precisión al detectar las opiniones negativas. 83,9%
- iii. Recall al detectar las opiniones positivas. 83,1%
- iv. Recall al detectar las opiniones negativas. 88,2%
- v. Precisión promedio. 85,7%
- vi. Recall promedio. 85,7%

Con una mejora aproximada de 3 decimas.

3. Búsqueda basada en palabras clave

4. Conclusiones