

Caso de estudio-Análisis de Sentimientos

Yoan Gutiérrez Vázquez
ygutierrez@dlsi.ua.es

PROPUESTA DE CASO DE ESTUDIO: TIPOS DE ANÁLISIS

- **Análisis del discurso:** “¡Ha sido una película emocionante y nos lo hemos pasado bien! No obstante, el servicio fue malo.”  Positivo
- **Análisis de frase/oración:** “¡Ha sido una película emocionante y nos lo hemos pasado bien!”  Positivo
- **Análisis de aspectos** tanto a nivel de frase como a nivel de discurso: “¡Ha sido una **película** emocionante y nos lo hemos pasado bien! No obstante, el **servicio** fue malo.”  Positivo  Negativo

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS

- **Análisis del discurso:** “¡Ha sido una película *emocionante* y nos lo hemos pasado *bien*! No obstante, el servicio fue *malo*.”

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS

- **Análisis del discurso:** “¡Ha sido una película **emocionante** y nos lo hemos pasado **bien**! No obstante, el servicio fue **malo**.”
- **Recursos:**
 - **LP:** Listado de palabras clasificadas como **positivas** {bueno, mejor, **emocionante**, **bien**, ...}
 - **LN:** Listado de palabras clasificadas como **negativas** {malo, pésimo, regular, **horrible**, ...}
- **Función heurística (ejemplo):**
 - *Positivo* $\sum \text{Palabras positivas} > \sum \text{Palabras negativas}$
 - *Negativo* $\sum \text{Palabras positivas} < \sum \text{Palabras negativas}$
 - *Neutral en otro caso*

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. **NEGADORES**

- **Análisis del discurso:** “¡Ha sido una película emocionante y nos lo hemos pasado bien! No obstante, el servicio no fue bueno.”



malo

- **Función Heurística (ejemplo):**
 - *Positivo* $\sum \text{Palabras positivas} > \sum \text{Palabras negativas}$
 - *Nagativo* $\sum \text{Palabras positivas} < \sum \text{Palabras negativas}$
 - *Neutral en otro caso*

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. NEGADORES (INVERSORES)

- **Recursos:**

- **LP:** Listado de palabras clasificadas como positivas {bueno, mejor, emocionante, bien, ...}
- **LN:** Listado de palabras clasificadas como negativas {malo, pésimo, regular, horrible, ...}
- **Linv:** Listado de palabras que invierten el sentido {no, ni, nunca, jamás, tampoco, nadie, nada, ningún, ninguno, ninguna , ...}

- **Preámbulo:**

- Palabra(término) **positiva**: $\begin{cases} \in LP \\ \in LN \text{ \& Modificada por otra palabra } \in L_{inv} \end{cases}$

no fue malo

- Palabra(término) **negativa**: $\begin{cases} \in LN \\ \in LP \text{ \& Modificada por otra palabra } \in L_{inv} \end{cases}$

no fue bueno

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. NEGADORES (INVERSORES)

- **Recursos:**

- **LP:** Listado de palabras clasificadas como positivas {bueno, mejor, emocionante, bien, ...}
- **LN:** Listado de palabras clasificadas como negativas {malo, pésimo, regular, horrible, ...}
- **Linv:** Listado de palabras que invierten el sentido {no, ni, nunca, jamás, tampoco, nadie, nada, ningún, ninguno, ninguna , ...}

- **Preámbulo:**

- Palabra(término) positiva:
 - $\in$ LP
 - $\in$ LN & Modificada por otra palabra $\in$ Linv
- Palabra(término) negativo:
 - $\in$ LN
 - $\in$ LP & Modificada por otra palabra $\in$ Linv

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. NEGADORES (INVERSORES)

- Recuerda

RETOS:

- **Doble negación:** no siempre anula el sentido de la negación
 - “**No** me **gustó nada**”
- Preámbulo
 - **Modificadores multi-palabras:** gestionar correctamente
 - {*Nunca jamás, Ni si quiera, ...*}
 - **Referencia:** ¿Cómo saber a qué palabra modifica?
 - “**el servicio no fue bueno**”

PROPOSTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. **MODIFICADORES AUMENTATIVOS (MA)/DIMINUTIVOS(MD)**

- **Recursos:**

- **LP:** Listado de palabras clasificadas como positivas {bueno, mejor, emocionante, bien, ...}
- **LN:** Listado de palabras clasificadas como negativas {malo, pésimo, regular, horrible, ...}
- **Lin_v:** Listado de palabras que invierten el sentido {no, ni, nunca, jamás, tampoco, nadie, nada, ningún, ninguno, ninguna , ...}
- **MA:** aumentativos {muy, mucho,...}
- **MD:** diminutivos {poco, menos, ...}

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. MODIFICADORES AUMENTATIVOS (MA)/DIMINUTIVOS(MD)

- Preámbulo:

- Palabra(término) **positiva**: $\ast \begin{cases} \frac{1}{\partial}, \text{Modificada por otra palabra} \in MD \\ \partial, \text{Modificada por otra palabra} \in MA \end{cases} , \partial > 1$

- Palabra(término) **negativa**: $\ast \begin{cases} \frac{1}{\partial}, \text{Modificada por otra palabra} \in MD \\ \partial, \text{Modificada por otra palabra} \in MA \end{cases} , \partial > 1$

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. MODIFICADORES AUMENTATIVOS (MA)/DIMINUTIVOS(MD)

- Preámbulo

Modificadores:

- **MA:** Adverbios (muy, mucho): "...fue **muy** bueno."
- **MD:** Adverbios (poco, algo ...): "...fue **poco** bueno."
- **MA:** Superlativos (**riquísimo**): "...fue **buenísimo**."
- **MD:** Diminutivos
- **MA:** Aumentativos
- ...

PROPUESTA DE CASO DE ESTUDIO: **RETOS**

- **Cómo determinar la **relación** entre las palabras subjetivas **claves** y:**
- **modificadores o negadores,**
- **o aspectos?**

The slide features decorative circuit-like lines in light blue and teal. These lines are composed of straight segments connected by small circles, resembling a stylized electronic circuit board. They are positioned along the left and right margins of the slide, framing the central text.

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS: **RELACIONAR ELEMENTOS**

Retos:

- Estrategias más comunes:
 - **Ventana de palabras**
 - Árboles sintácticos
 - Árboles de dependencia
 - Otros...

PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: N-gramas

“*¡Ha sido una película emocionante y nos lo hemos pasado muy bien!*”

1-grama (1P, 1Mod)	2-grama (1 Mod, 1P, 1Mod+1P)	3-grama (1P, 1Mod+1P,...)	N-grama $\partial = 1,1$
...	...	...	...
hemos	lo hemos	nos lo hemos	...
pasado	hemos pasado	lo hemos pasado	...
muy	pasado muy	hemos pasado muy	...
bien	muy bien	pasado muy bien	...

$$1$$

$$(1 * 1, 1) = 1, 1$$

$$(1 * 1, 1) = 1, 1$$

PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: N-gramas

“No obstante, el servicio no fue bueno.”

1-grama (1P, 1Neg)	2-grama (1Neg, 1P)	3-grama (2Neg, 1Neg+1P,...)	N-grama	$\partial = 1,1$ $\alpha = -1$
...	...	...	...	
el	obstante el	no obstante, el	...	
servicio	el servicio	obstante el servicio	...	
no	servicio no	el servicio no	...	
fue	no fue	servicio no fue	...	
bueno	fue bueno	<u>no fue bueno</u>	...	

$$(-1 * 0) + 1 = 1$$

$$(-1 * 0) + (-1 * 0) + 1 = 1$$

$$(-1 * 0) + (-1 * 0) + (-1 * 1) = -1$$

Sentido
contrario

O decidir
anularlo $\alpha = 0$

PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: Skip-gramas (k -skip- n -gram)

Formalmente,

- **n -gramas** es una **subsecuencia consecutiva** de longitud n de alguna secuencia de tokens $w_1 \dots w_n$.
- **k -skip- n -gramas** es una **subsecuencia** de longitud n donde los **componentes ocurren** a una **distancia máxima** k entre sí.

PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: Skip-gramas (k -skip- n -gram)

“¡Ha sido una película emocionante y nos lo hemos pasado muy bien!”

1-skip-2-gramas	1-skip-3-gramas	2-skip-2-gramas	k-skip-n-gramas
...	...	...	...
ha sido	ha sido una	ha sido	...
ha una	ha una película	ha una	...
sido una	ha sido película	ha película	...
sido película	sido una película	sido una	...
una película	sido película emocionante	sido película	...
...	...	...	...

PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: 3-gramas, **discurso**

+1

$\partial = 1,1$

+1

“¡Ha sido una película emocionante y nos lo hemos pasado muy bien! No obstante, el servicio no fue bueno.”

-1

2,1

-1



PROPUESTA DE CASO DE ESTUDIO: RETOS

Ventana de palabras: 3-gramas, **aspectos**

*“¡Ha sido una **película** emocionante y nos lo hemos pasado muy bien! No obstante, el **servicio** no fue bueno.”*

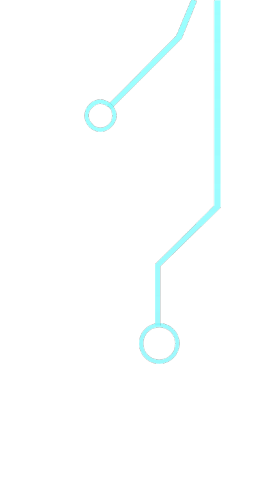
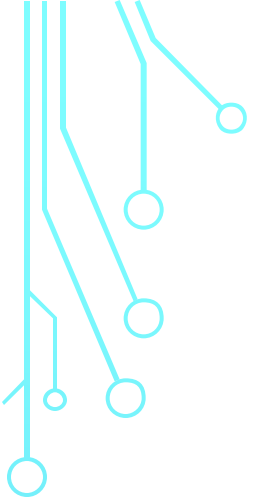


obstante, el **servicio** no fue **bueno**



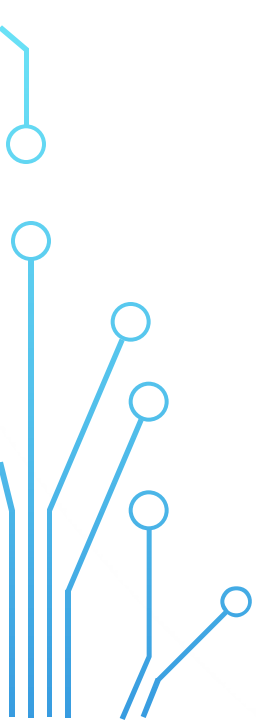
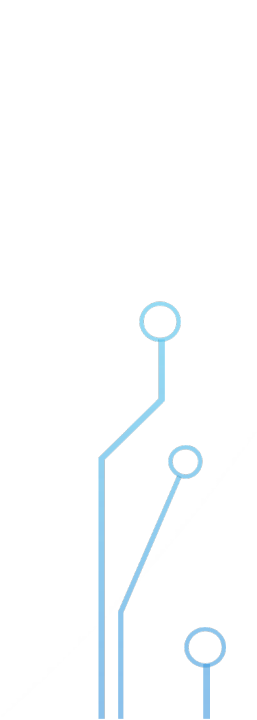
sido una **película** emocionante y





PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS: **RELACIONAR ELEMENTOS**

Retos

- Estrategias más comunes:
 - Ventana de palabras
 - Árboles sintácticos
 - **Árboles de dependencia**
 - Otros...
- 
- 

PROPUESTA DE CASO DE ESTUDIO: RETOS

Arboles de dependencia:

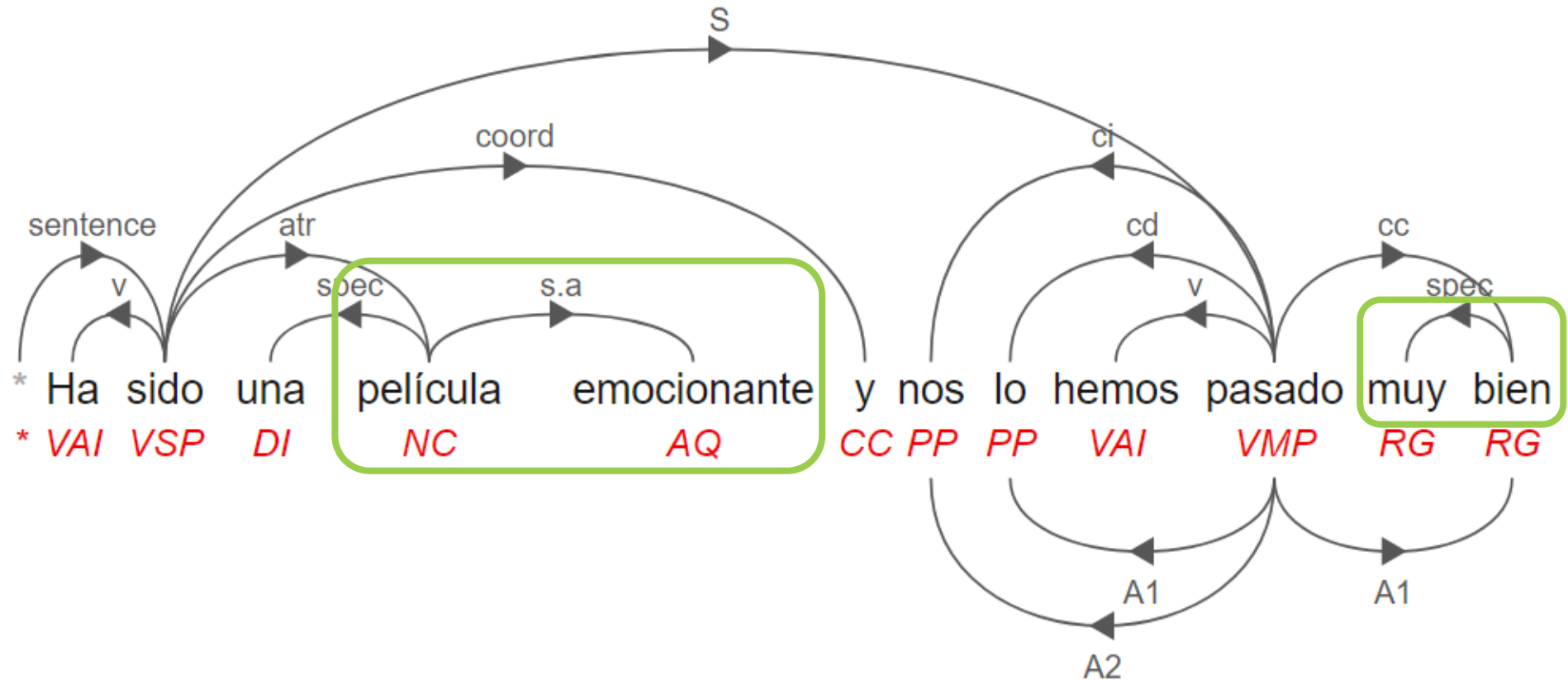
“¡Ha sido una película emocionante y nos lo hemos pasado muy bien! No obstante, el servicio no fue bueno.”

A continuación veremos dos ejemplos utilizando la demo online de Freeling

<http://nlp.lsi.upc.edu/freeling/demo/demo.php>

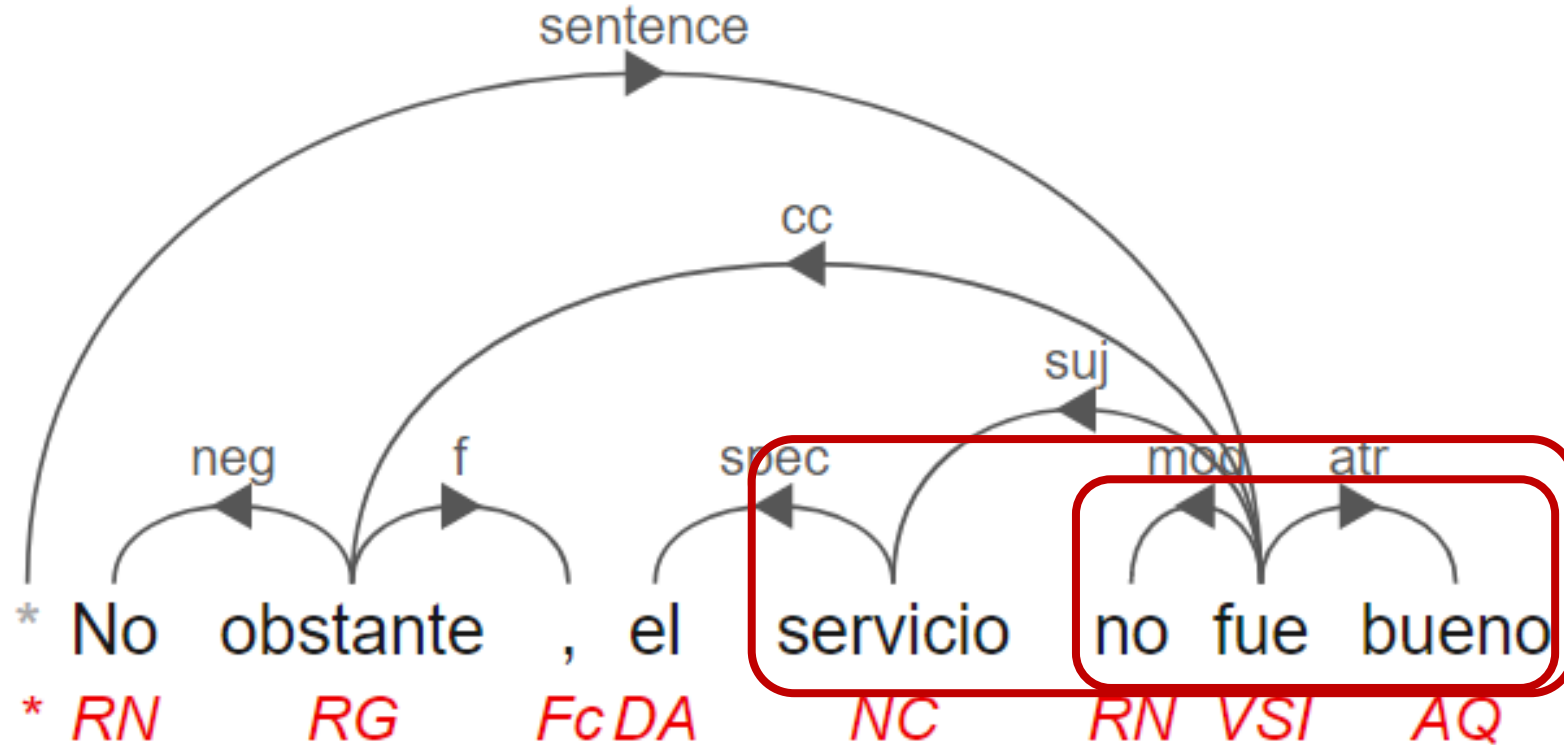
PROPUESTA DE CASO DE ESTUDIO: RETOS

Arboles de dependencia: “jHa sido una película emocionante y nos lo hemos pasado muy bien! bien!”



PROPUESTA DE CASO DE ESTUDIO: RETOS

Arboles de dependencia: No obstante, el servicio no fue bueno.



PROPUESTA DE CASO DE ESTUDIO: USO DE **APRENDIZAJE AUTOMÁTICO**



Problem Solving with Machine Learning

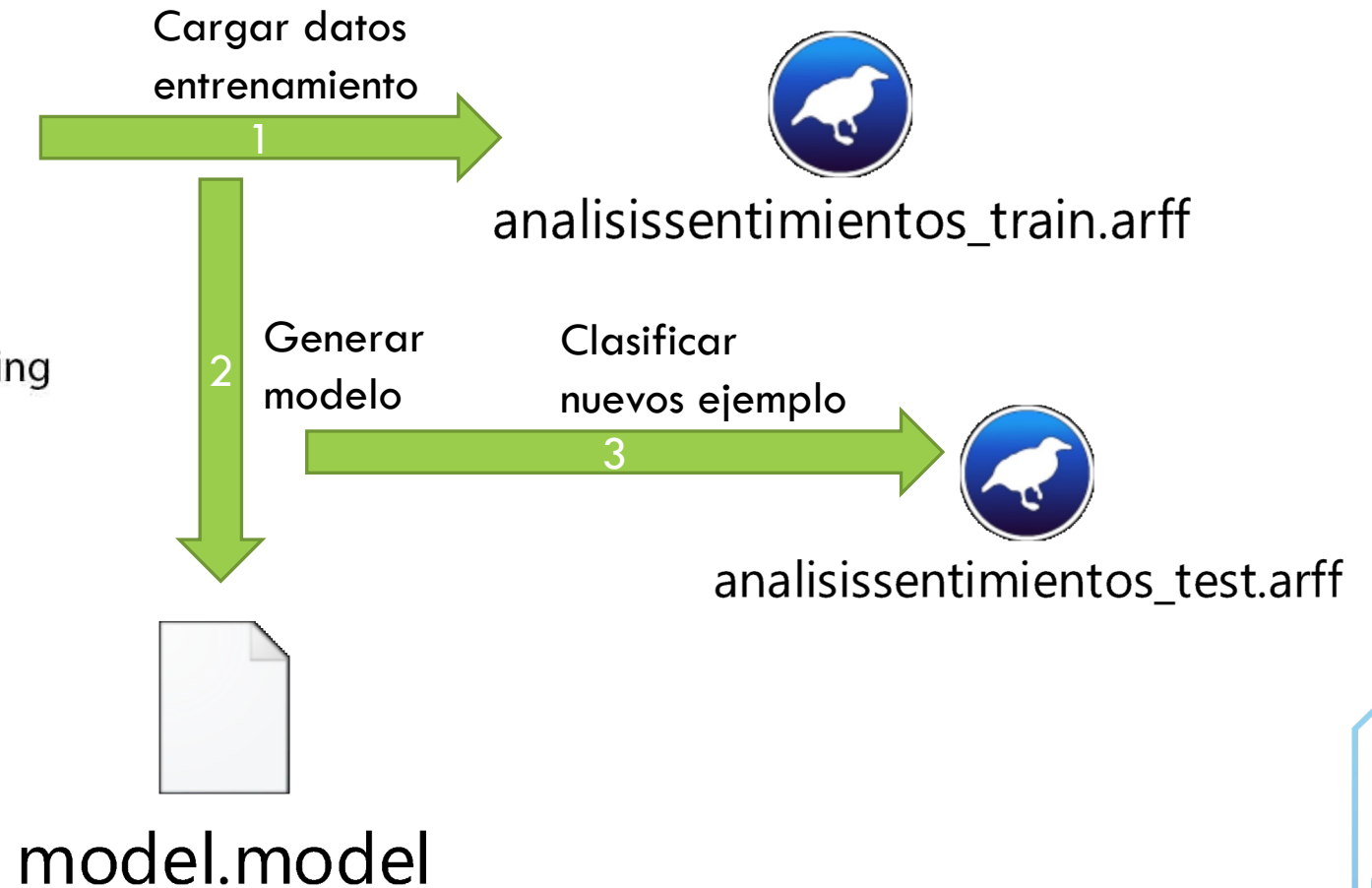
VS

Función Heurística

PROPUESTA DE CASO DE ESTUDIO: BOLSA DE PALABRAS. **FUNCION HEURISTICA**

- **Función Heurística (ejemplo):**
 - *Positivo* $\sum \text{Palabras positivas} > \sum \text{Palabras negativas}$
 - *Nagativo* $\sum \text{Palabras positivas} < \sum \text{Palabras negativas}$
 - *Neutral en otro caso*

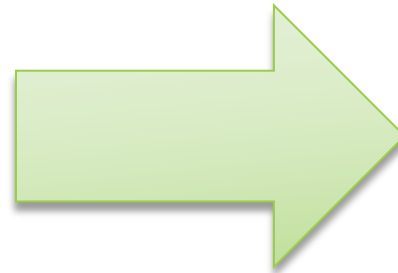
PROPUESTA DE CASO DE ESTUDIO: USO DE **APRENDIZAJE AUTOMÁTICO**. ARFF WEKA



[illegible][illegible]

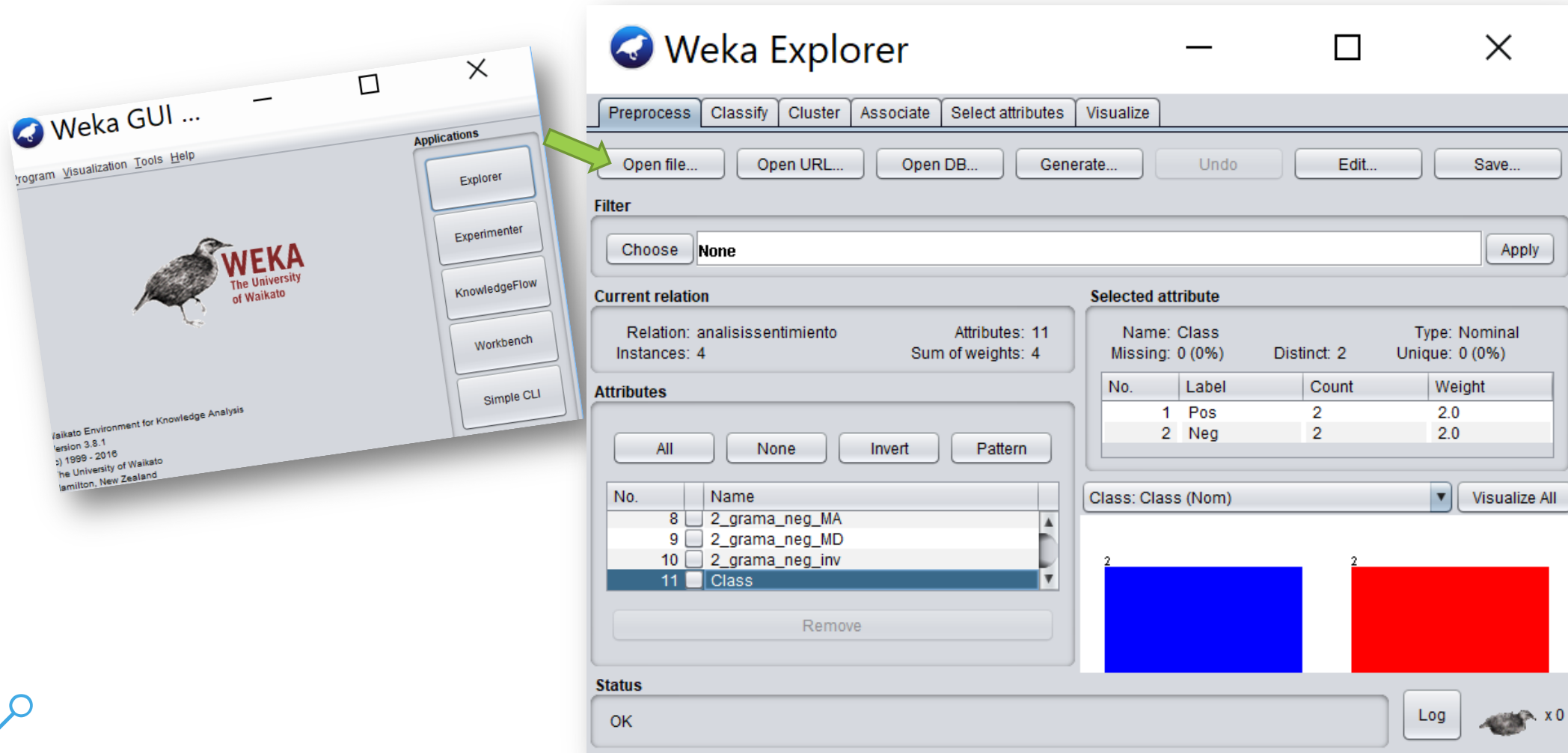
PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. WEKA

```
@relation analisis sentimientos
@attribute 1_gram_pos numeric
@attribute 2_gram_pos numeric
@attribute 2_gram_pos_MA numeric
@attribute 2_gram_pos_MD numeric
@attribute 2_gram_pos_inv numeric
@attribute 1_gram_neg numeric
@attribute 2_gram_neg numeric
@attribute 2_gram_neg_MA numeric
@attribute 2_gram_neg_MD numeric
@attribute 2_gram_neg_inv numeric
@attribute 'Class' {Pos,Neg}
@data
1,1,0,0,0,0,0,0,0,0,Pos
1,1,1,0,0,1,1,0,0,0,Pos
0,0,0,0,0,1,1,0,0,0,Neg
2,0,0,0,2,0,0,0,0,0,Neg
....
```



analisis sentimientos_train.arff

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **ENTRENAR MODELO**



The image displays two windows from the Weka machine learning software. The 'Weka GUI' window on the left shows the 'Applications' sidebar with 'Explorer' selected. A green arrow points from 'Explorer' to the 'Weka Explorer' window on the right.

The 'Weka Explorer' window shows the 'Preprocess' tab. The 'Filter' section has 'Choose' selected and 'None' in the dropdown. The 'Current relation' section shows 'Relation: analisis sentiment' with 11 attributes and 4 instances. The 'Attributes' list includes '2_grama_neg_MA', '2_grama_neg_MD', '2_grama_neg_inv', and 'Class'. The 'Selected attribute' section shows 'Name: Class' with a 'Type: Nominal' and a table of counts for 'Pos' and 'Neg' classes. The 'Class: Class (Nom)' dropdown is set to 'Class (Nom)'. The 'Visualize All' button is visible. The 'Status' bar at the bottom shows 'OK'.

Weka Explorer - Filter Section

Choose: None [Apply]

Weka Explorer - Current relation Section

Relation: analisis sentiment Attributes: 11
Instances: 4 Sum of weights: 4

Weka Explorer - Attributes Section

All None Invert Pattern

No.	Name
8	2_grama_neg_MA
9	2_grama_neg_MD
10	2_grama_neg_inv
11	Class

Remove

Weka Explorer - Selected attribute Section

Name: Class
Missing: 0 (0%) Distinct: 2 Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Pos	2	2.0
2	Neg	2	2.0

Class: Class (Nom) [Visualize All]

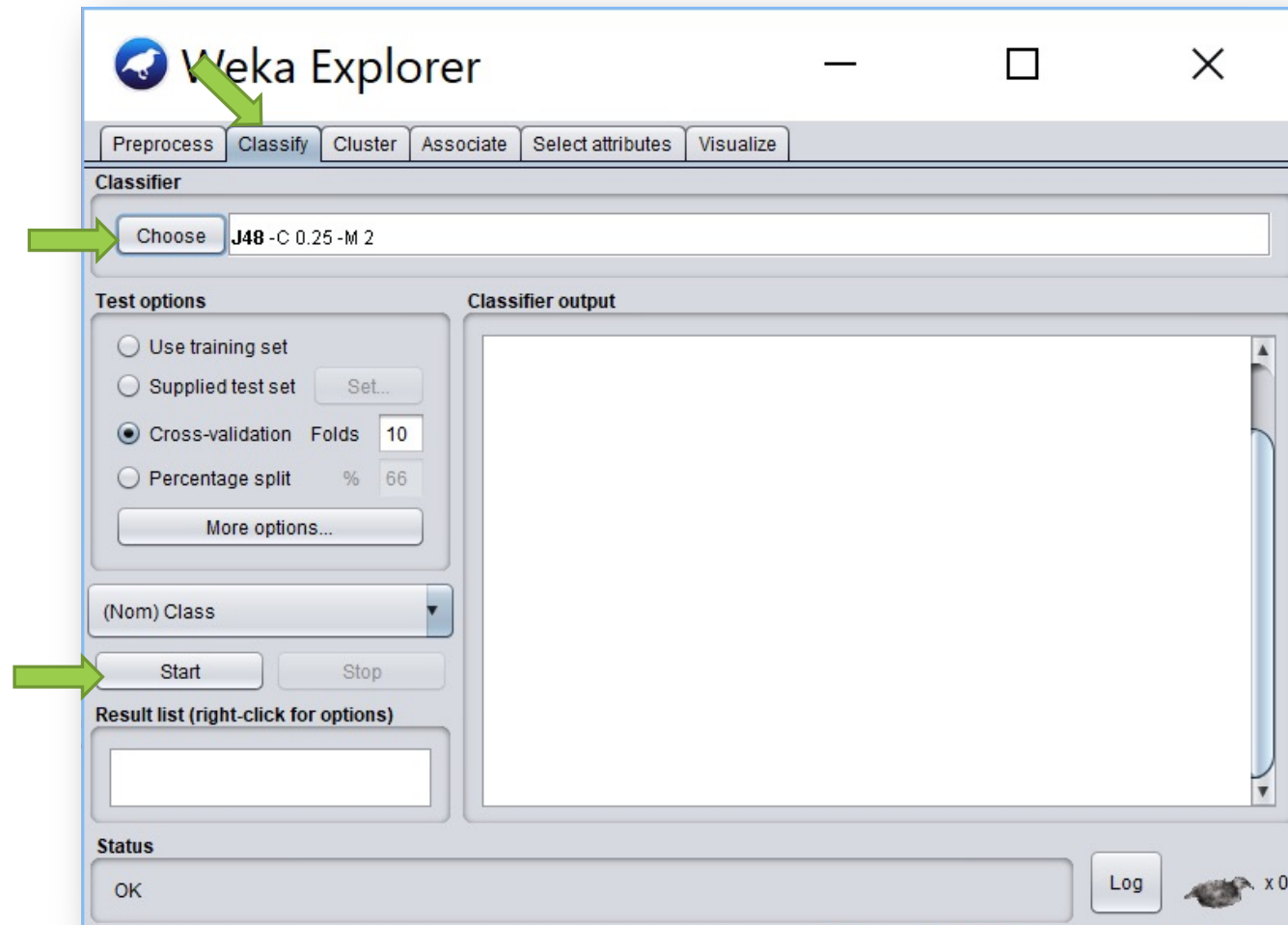
Weka Explorer - Visualize Section

2 [Blue Bar] 2 [Red Bar]

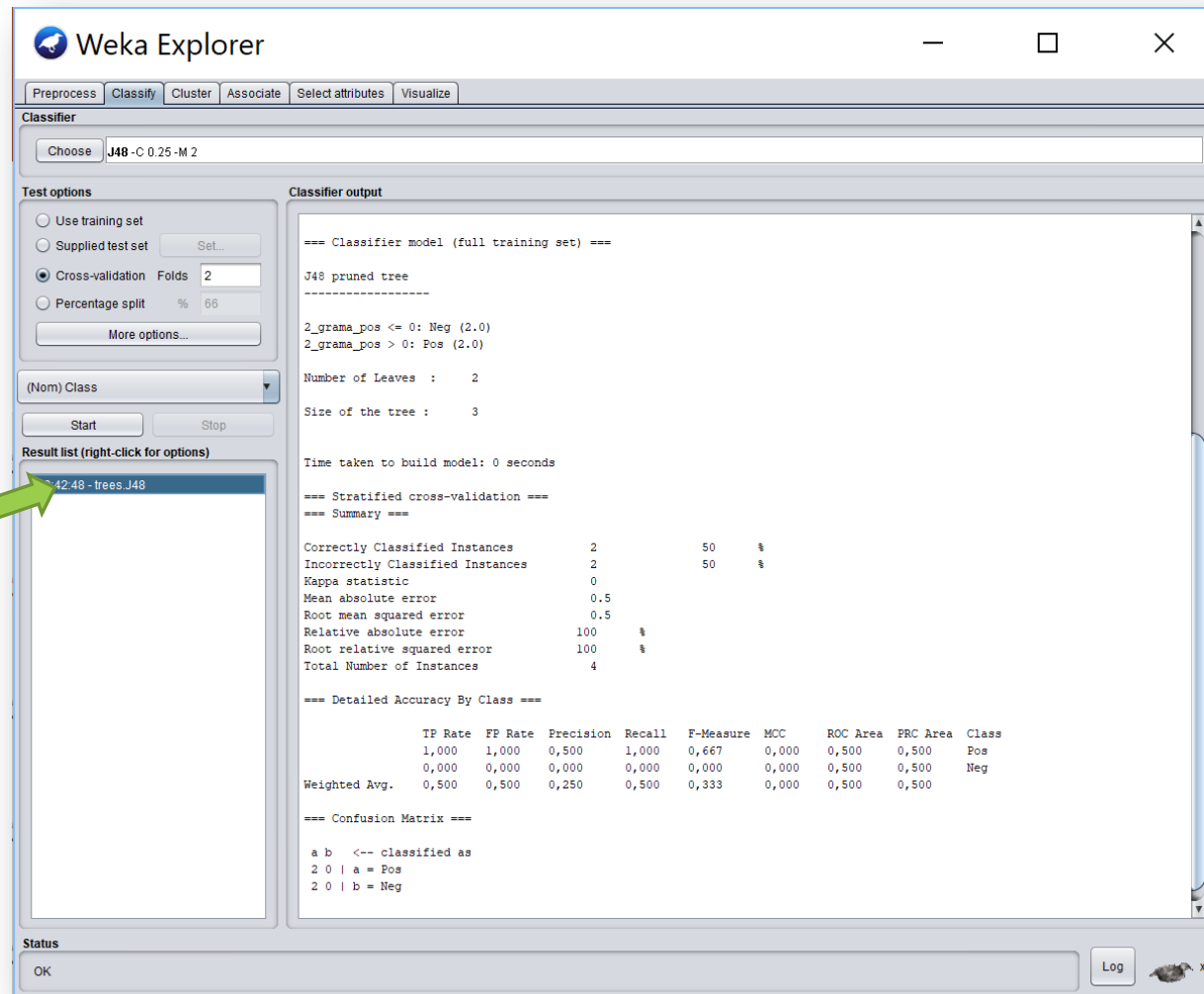
Weka Explorer - Status Section

OK [Log] x 0

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **ENTRENAR MODELO**



PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. ENTRENAR MODELO



The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected. The classifier chosen is 'J48 - C 0.25 - M 2'. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 2. The 'Result list' on the left shows a single entry '42.48 - trees.J48' highlighted with a green arrow. The 'Classifier output' pane displays the following results:

```
=== Classifier model (full training set) ===  
J48 pruned tree  
-----  
2_grama_pos <= 0: Neg (2.0)  
2_grama_pos > 0: Pos (2.0)  
  
Number of Leaves :    2  
Size of the tree :    3  
  
Time taken to build model: 0 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
  
Correctly Classified Instances      2          50 %  
Incorrectly Classified Instances    2          50 %  
Kappa statistic                     0  
Mean absolute error                 0.5  
Root mean squared error             0.5  
Relative absolute error             100 %  
Root relative squared error         100 %  
Total Number of Instances          4  
  
=== Detailed Accuracy By Class ===  
  
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class  
      1,000    1,000    0,500    1,000    0,667    0,000    0,500    0,500    Pos  
      0,000    0,000    0,000    0,000    0,000    0,000    0,500    0,500    Neg  
Weighted Avg.   0,500    0,500    0,250    0,500    0,333    0,000    0,500    0,500  
  
=== Confusion Matrix ===  
  
a b  <-- classified as  
2 0 | a = Pos  
2 0 | b = Neg
```

The 'Status' bar at the bottom shows 'OK' and a 'Log' button.

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. ENTRENAR MODELO

Scheme: **weka.classifiers.trees.J48 -C 0.25**

-M 2

Test: training set

J48 pruned tree

2_grama_pos <= 0: Neg (2.0)

2_grama_pos > 0: Pos (2.0)

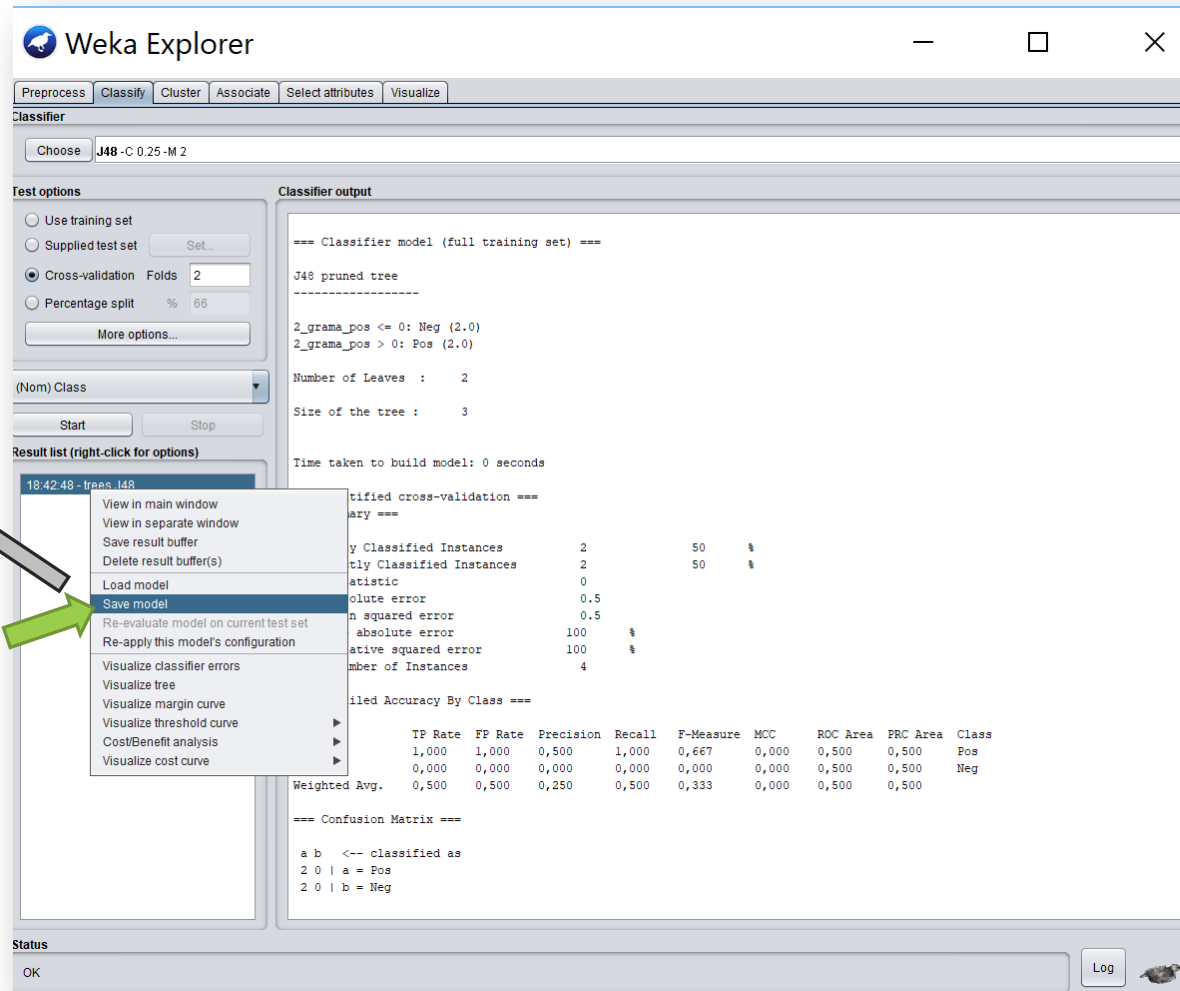
=== Summary ===

Correctly Classified Instances	4	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
K&B Relative Info Score	400	%	
K&B Information Score	4	bits	1
bits/instance			
Class complexity order 0	4	bits	1
bits/instance			
Class complexity scheme	0	bits	0
bits/instance			
Complexity improvement (Sf)	4	bits	1
bits/instance			
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	4		

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **GUARDAR MODELO**

model.model

SAVE



PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. MODELO WEKA

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Test: training set

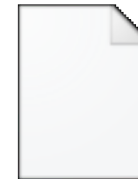
J48 pruned tree

2_grama_pos <= 0: Neg (2.0)

2_grama_pos > 0: Pos (2.0)



Problem Solving with Machine Learning

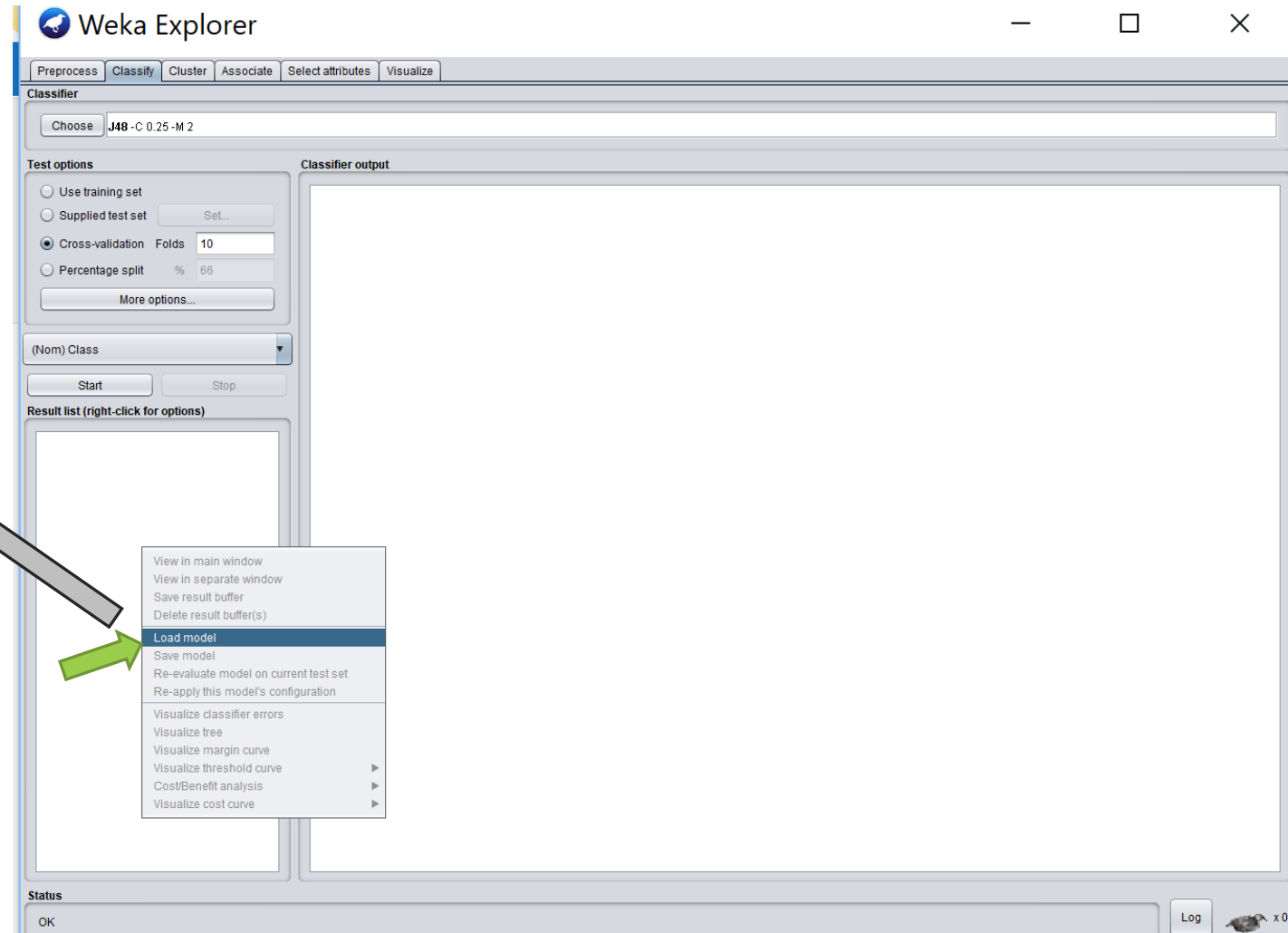


model.model

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **CARGAR MODELO**


model.model

LOAD



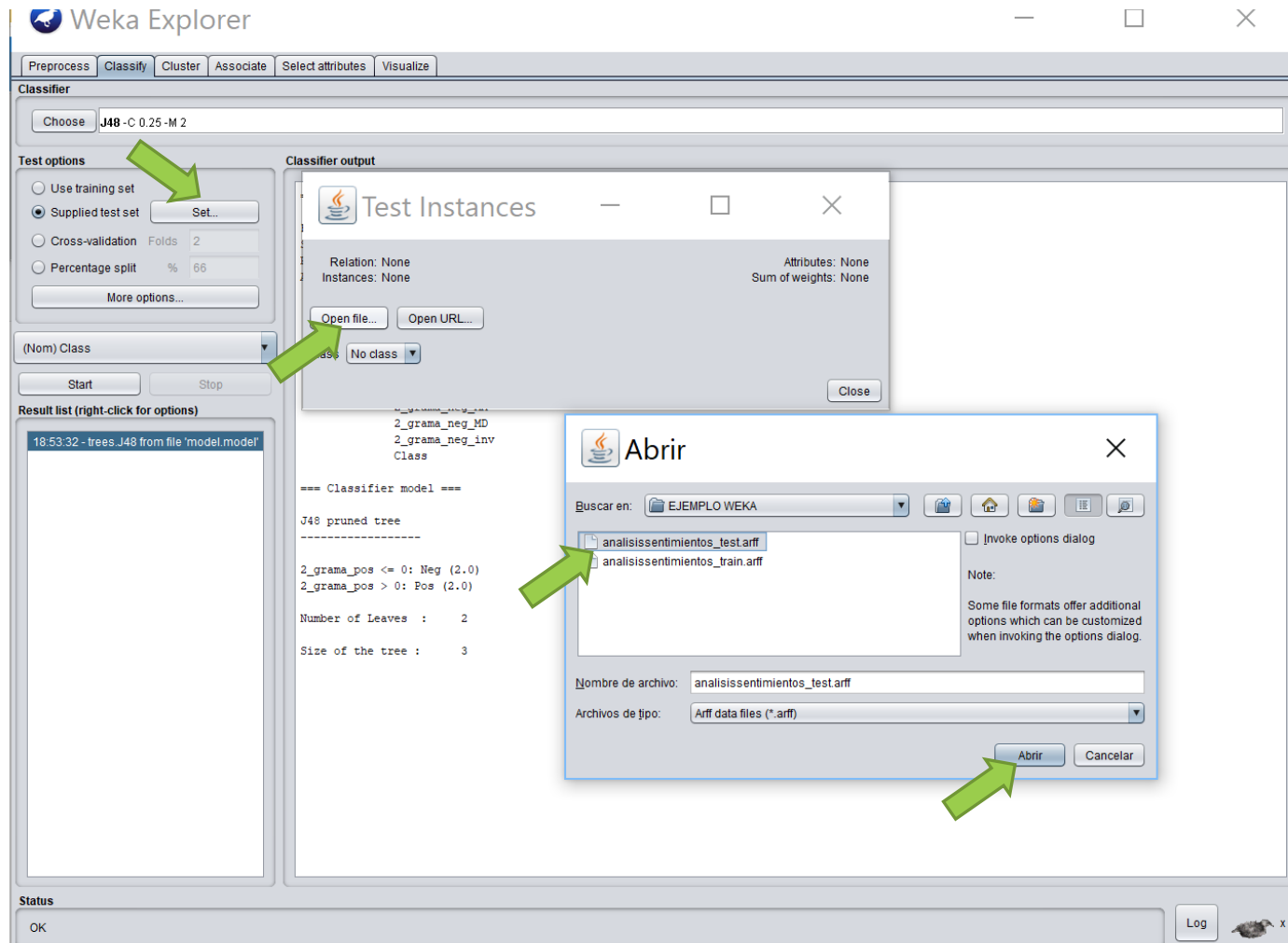
PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **CLASIFICAR NUEVOS EJEMPLOS**

```
@relation analisis sentimentos
@attribute 1_gram_pos numeric
@attribute 2_grama_pos numeric
@attribute 2_grama_pos_MA numeric
@attribute 2_grama_pos_MD numeric
@attribute 2_grama_pos_inv numeric
@attribute 1_gram_neg numeric
@attribute 2_grama_neg numeric
@attribute 2_grama_neg_MA numeric
@attribute 2_grama_neg_MD numeric
@attribute 2_grama_neg_inv numeric
@attribute 'Class' {Pos,Neg}
@data
1,1,0,0,0,0,0,0,0,0,?
1,1,1,0,0,1,1,0,0,0,?
0,0,0,0,0,1,1,0,0,0,?
2,0,0,0,2,0,0,0,0,0,?
```

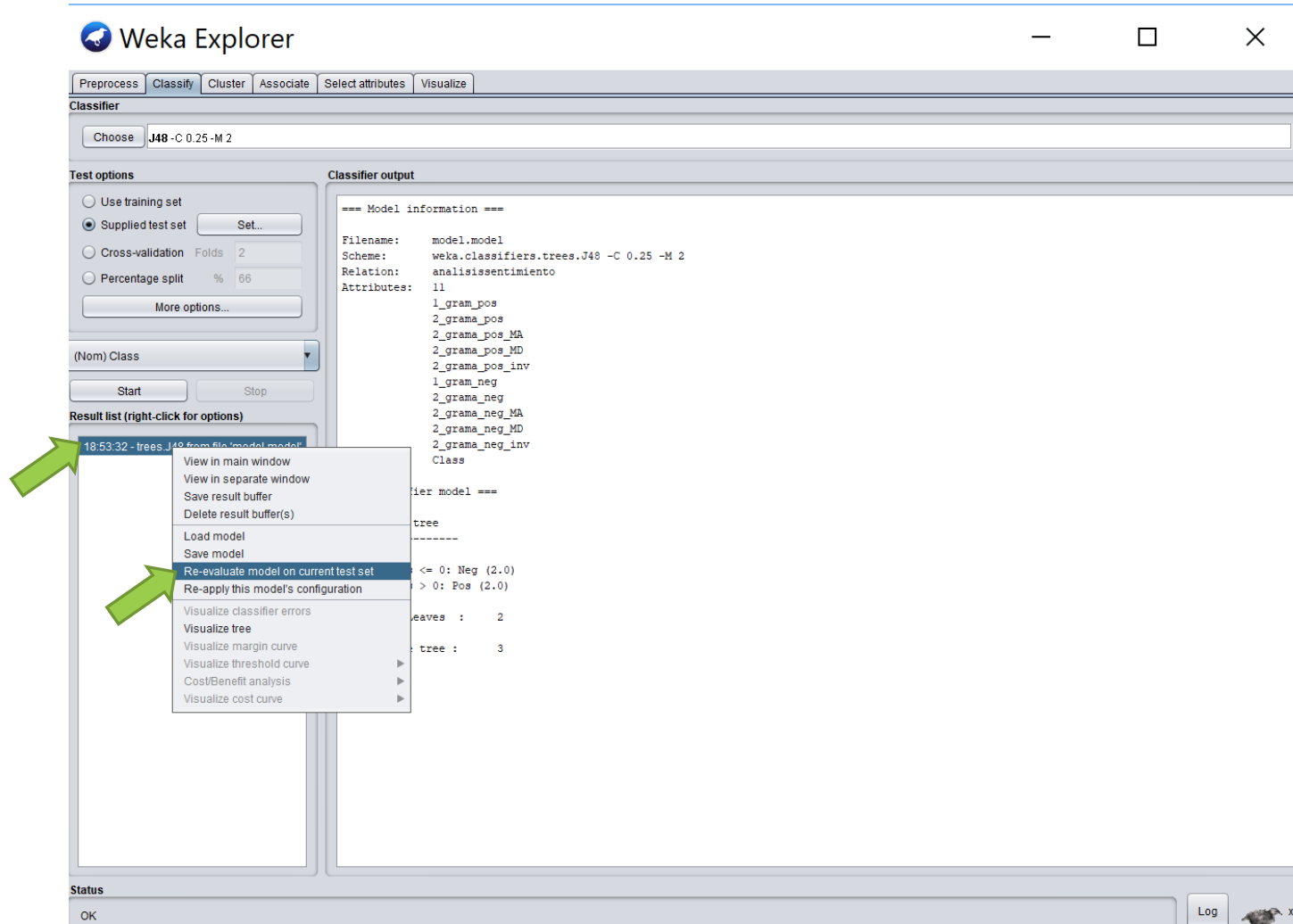


analisis sentimentos_test.arff

PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. **EVALUAR NEVOS EJEMPLOS**



PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. EVALUAR NEVOS EJEMPLOS



PROPUESTA DE CASO DE ESTUDIO: USO DE APRENDIZAJE AUTOMÁTICO. CLASIFICAR NUEVOS EJEMPLOS

```
@relation analisis sentimentos
@attribute 1_gram_pos numeric
@attribute 2_gram_pos numeric
@attribute 2_gram_pos_MA numeric
@attribute 2_gram_pos_MD numeric
@attribute 2_gram_pos_inv numeric
@attribute 1_gram_neg numeric
@attribute 2_gram_neg numeric
@attribute 2_gram_neg_MA numeric
@attribute 2_gram_neg_MD numeric
@attribute 2_gram_neg_inv numeric
@attribute 'Class' {Pos,Neg}
@data
1,1,0,0,0,0,0,0,0,0,?
1,1,1,0,0,1,1,0,0,0,?
0,0,0,0,0,1,1,0,0,0,?
2,0,0,0,2,0,0,0,0,0,?
```



analisis sentimentos_test.arff

=== Predictions on user test set ===

inst#	actual	predicted
1	?	Pos
2	?	Pos
3	?	Neg
4	?	Neg

Caso de estudio-Análisis de Sentimientos

Yoan Gutiérrez Vázquez
ygutierrez@dlsi.ua.es