

Inteligencia de Negocio Trabajo #2

Integrantes



Gracia Cusme Jhon.
e0802268615@live.uleam.edu.ec



Mantuano Quijije Moises.
e1316435930@live.uleam.edu.ec



Laz Arteaga Anthony.
e1315603934@live.uleam.edu.ec

Docente:

Ing. Fabricio Javier Rivadeneira Zambrano

ALGORITMO DECISION TREES, CON WEKA



¿QUÉ ES WEKA?

Waikato Environment for Knowledge Analysis

Es una herramienta de minería de datos / aprendizaje automático desarrollada por el Departamento de Ciencias de la Computación de la Universidad de Waikato, Nueva Zelanda.

Weka también es un ave que se encuentra solo en las islas de Nueva Zelanda

PRINCIPALES CARACTERÍSTICAS

- 49 herramientas de preprocesamiento de datos
- 76 algoritmos de clasificación / regresión
- 8 algoritmos de agrupamiento
- 3 algoritmos para encontrar reglas de asociación
- 15 evaluadores de atributos / subconjuntos + 10 algoritmos de búsqueda para la selección de características



PREPROCESAMIENTO DE LOS DATOS



- Los datos se pueden importar desde un archivo en varios formatos: ARFF, CSV, C4.5, binario.
- Los datos también se pueden leer desde una URL o desde una base de datos SQL (usando JDBC).
- Las herramientas de preprocesamiento en WEKA se llaman "filtros".
- WEKA contiene filtros para:
- Discretización, normalización, remuestreo, selección de atributos, transformación y combinación de atributos

DECISION TREES O ÁRBOL DE DECISIÓN

- El algoritmo de árboles de decisión es un algoritmo de clasificación y regresión para el modelado de predicción de atributos discretos y continuos.
- Es uno de los enfoques de modelado predictivo utilizadas en estadísticas, minería de datos y aprendizaje automático. Los modelos de árbol, donde la variable de destino puede tomar un conjunto finito de valores se denominan árboles de clasificación. En estas estructuras de árbol, las hojas representan etiquetas de clase y las ramas representan las conjunciones de características que conducen a esas etiquetas de clase.

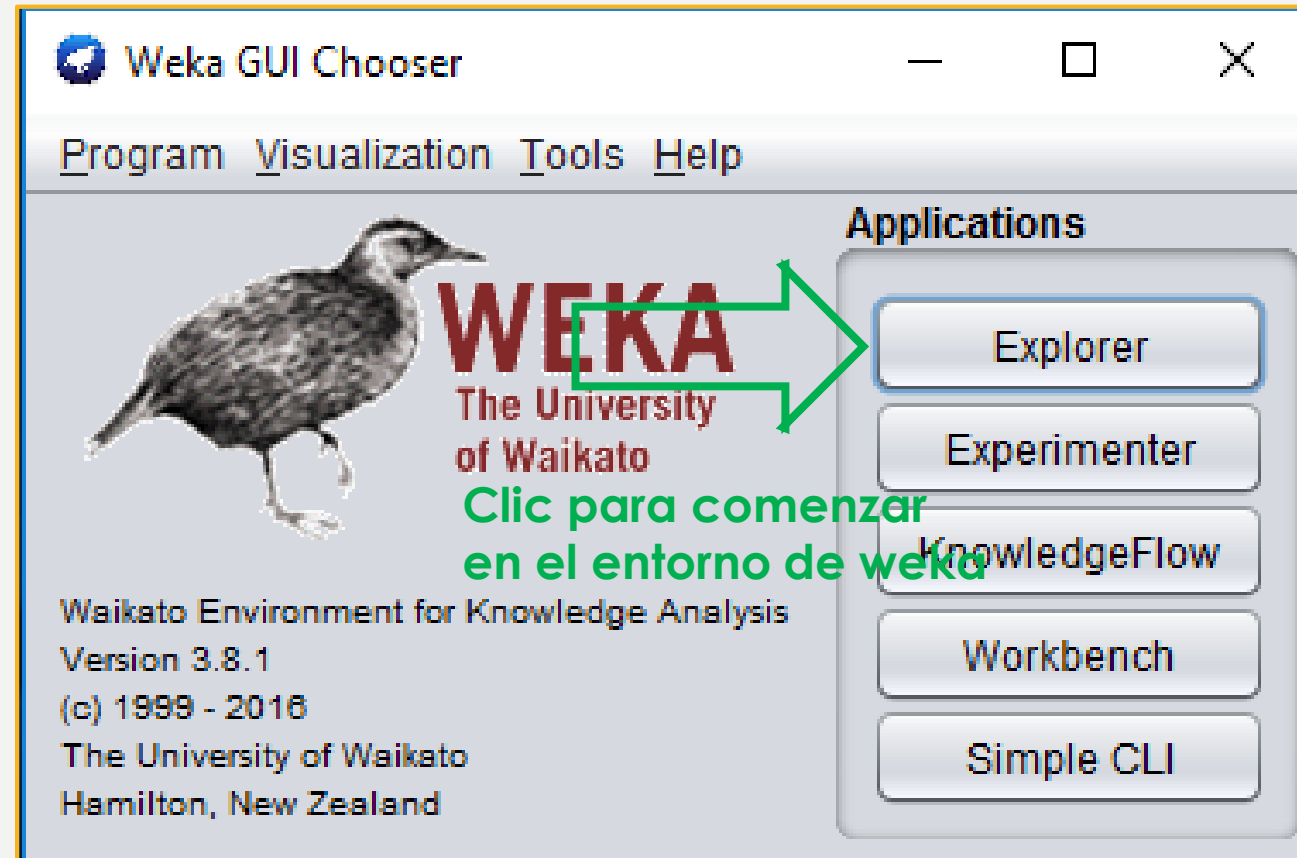
CLASIFICACIÓN.



Los árboles de decisión utilizados en la minería de datos son de dos tipos principales:

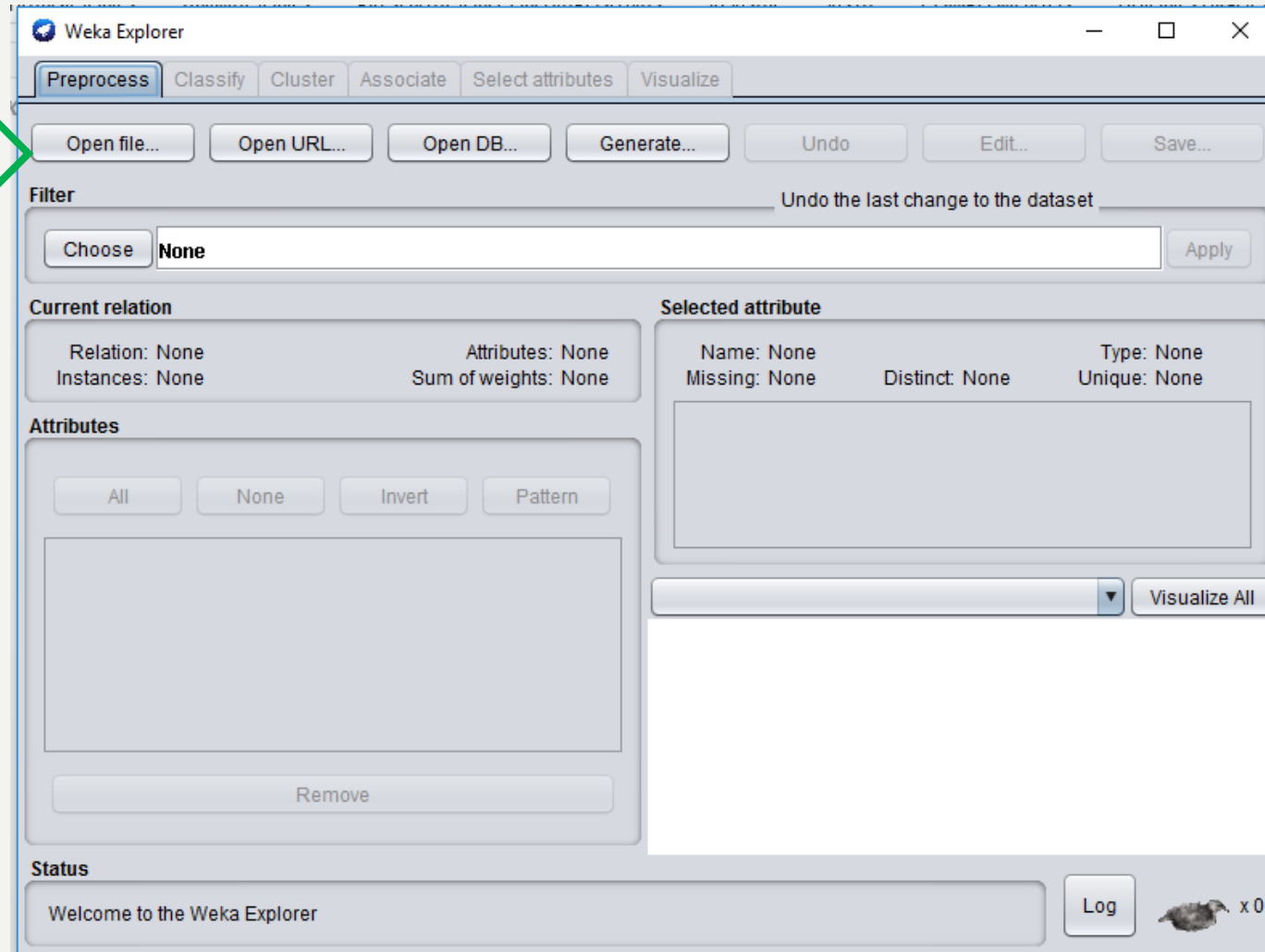
- **Árboles de clasificación** es cuando el resultado predicho es la clase a la que pertenecen los datos.
- **Árboles de regresión** es cuando el resultado predicho se puede considerar un número real (por ejemplo, el precio de una casa, o la longitud de un paciente de la estancia en un hospital).

Práctica



AÑADIR DATOS.

Clic para buscar los
datos guardados



Vamos a trabajar con unos datos autóctonos de weka “datos de los sobrevivientes del titanic”

FORMATO ARCHIVO.ARFF

Nombre:

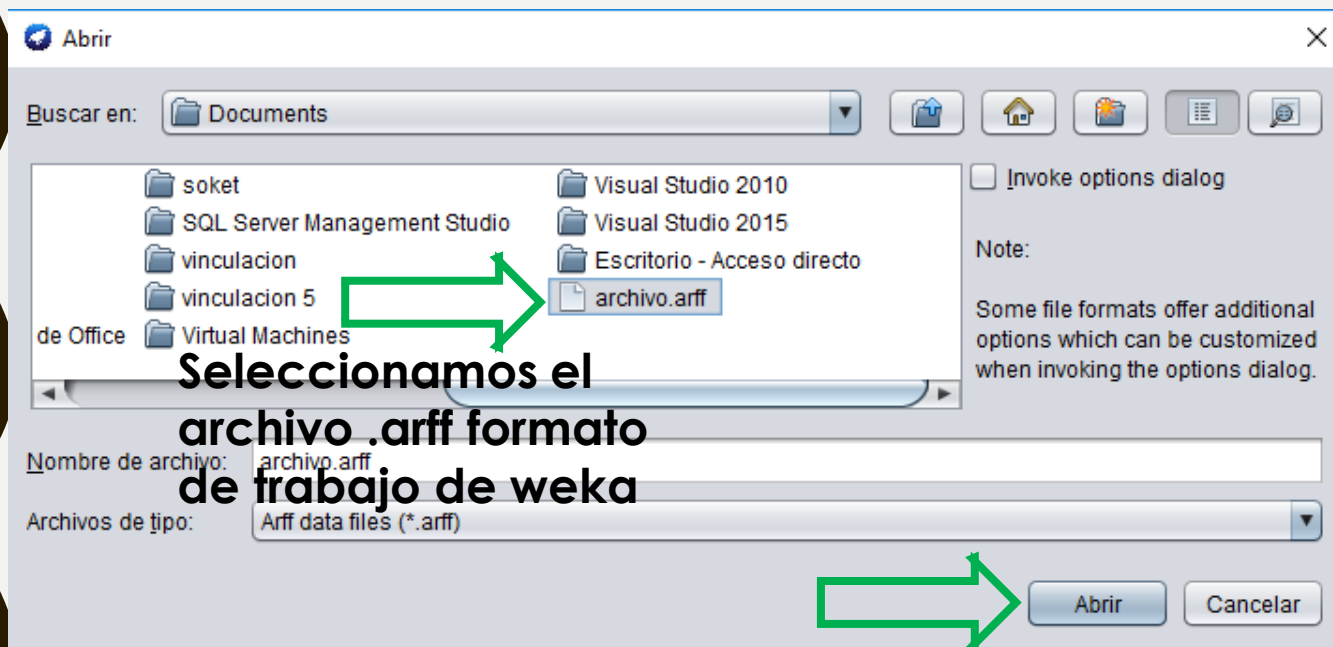
Atributos

Datos

```
Archivo  Edición  Formato  Ver  Ayuda
@RELATION relation
@ATTRIBUTE class {"1st","2nd","3rd","crew"}
@ATTRIBUTE age {"adult","child"}
@ATTRIBUTE sex {"male","female"}
@ATTRIBUTE survived {"yes","no"}
@DATA
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
1st,adult,male,yes
```

Que vamos a Analizar con los datos

- ¿Cuántas personas sobrevivieron en el titanic a que clase pertenecía si era hombre mujer o , adulto o niño ?

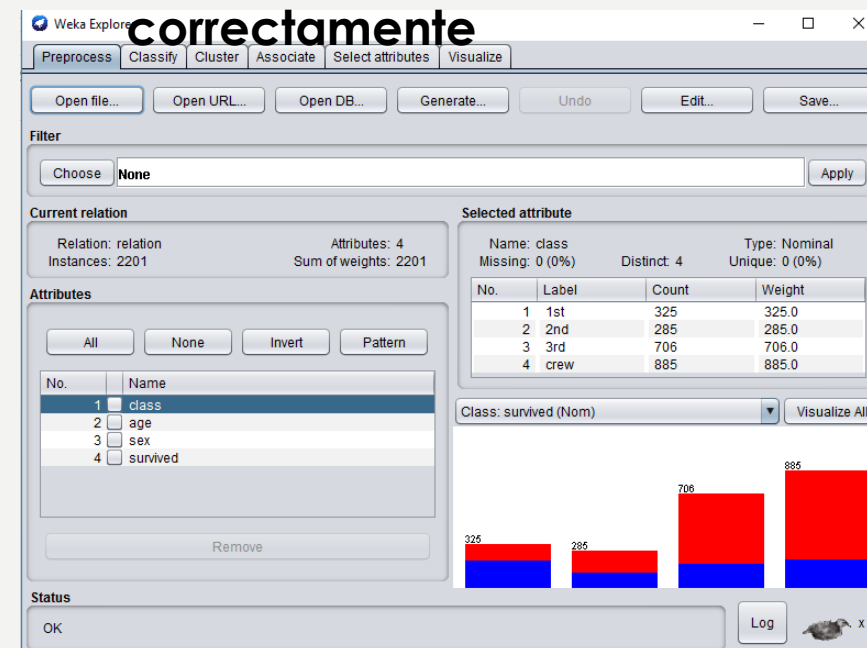


**Seleccionamos el
archivo .arff formato
de trabajo de weka**

**Seleccionamos el
archivo .arff formato
de trabajo de weka**

Datos ingresados

correctamente



Nombres
de los
Atributos



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: relation Instances: 2201 Attributes: 4 Sum of weights: 2201

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> class
2	<input type="checkbox"/> age
3	<input type="checkbox"/> sex
4	<input type="checkbox"/> survived

Remove

Selected attribute: Name: class Missing: 0 (0%) Distinct: 4 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	1st	325	325.0
2	2nd	285	285.0
3	3rd	706	706.0
4	crew	885	885.0

Class: survived (Nom) Visualize All

Status: OK Log x 0

Cantidad
de datos
dependien
do del
atributo en
este caso
“atributo
clase”

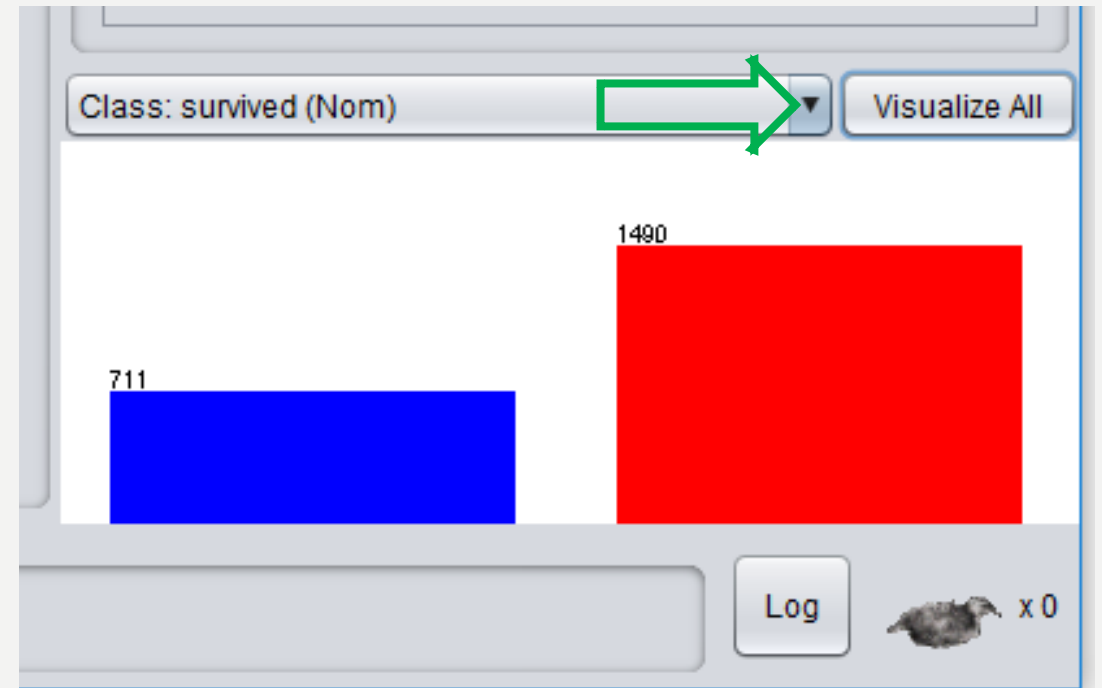
Mostrando la cantidad de datos por cada atributo



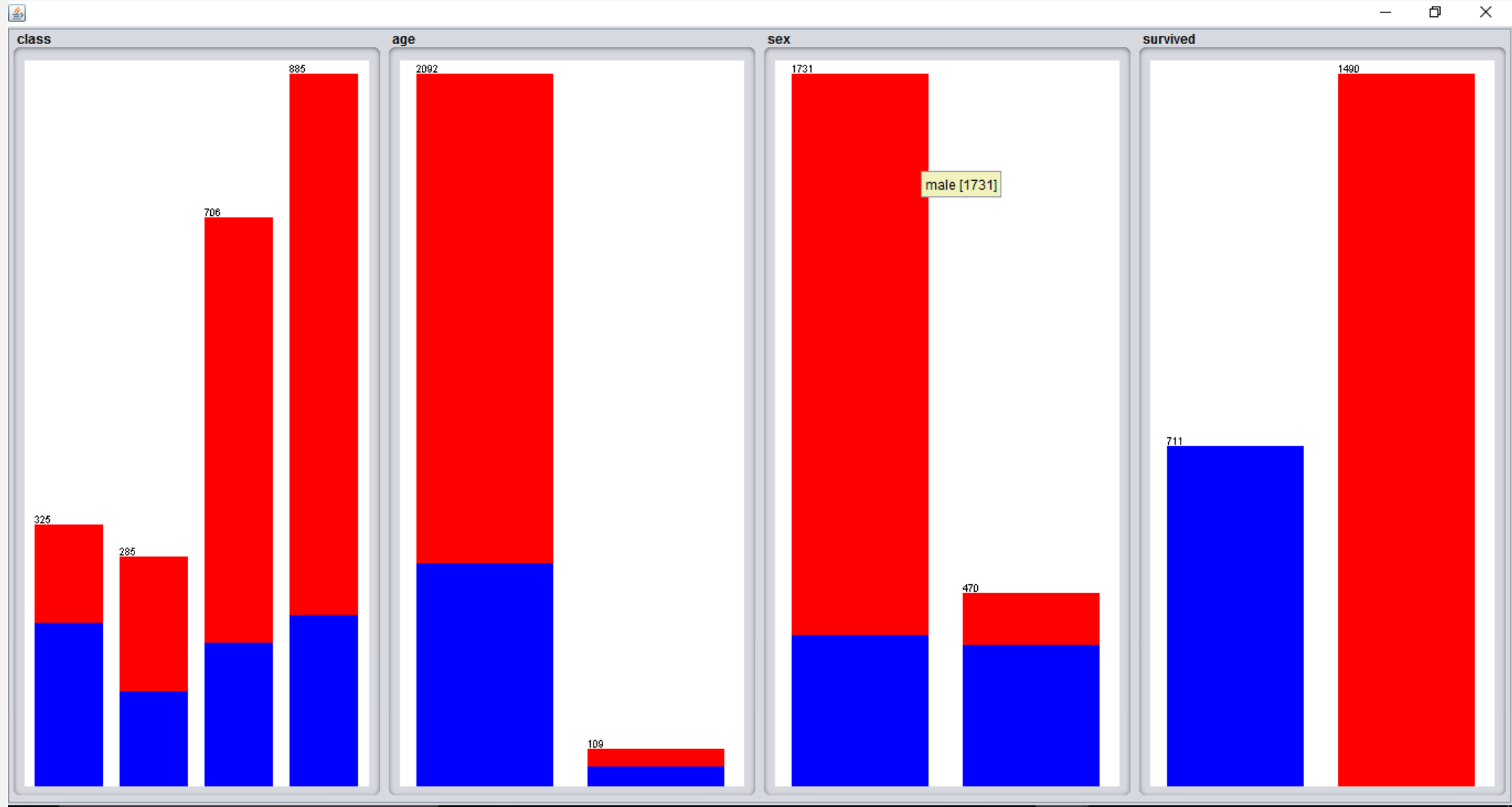
Seleccionamos todos los atributos.

All		None		Invert		Pattern	
No.	Name						
1	<input checked="" type="checkbox"/> class						
2	<input checked="" type="checkbox"/> age						
3	<input checked="" type="checkbox"/> sex						
4	<input checked="" type="checkbox"/> survived						

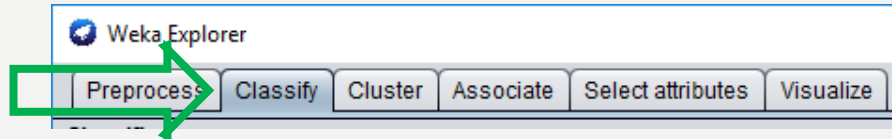
Clic para visualizar los datos



Datos de acuerdo a cada a tributo
clase,años,sexo,sobrevivio.

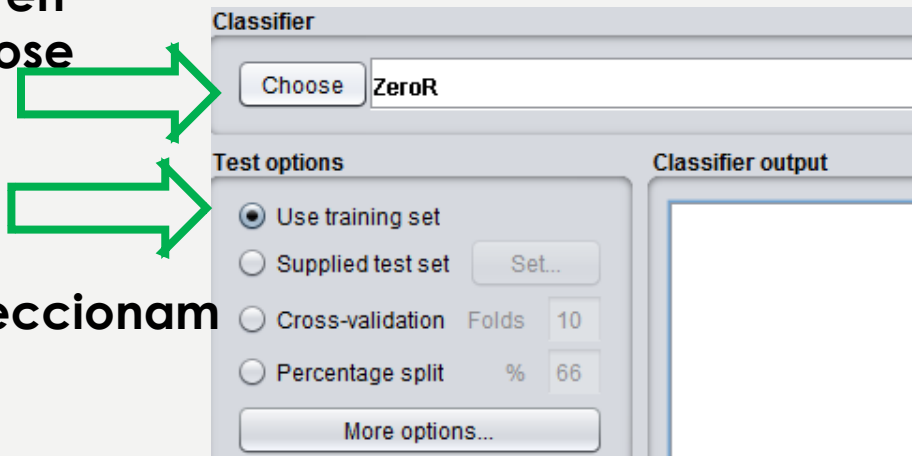


CREANDO NUESTRO ÁRBOL DE DECISIÓN

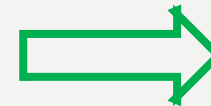


Clic en esta opción.

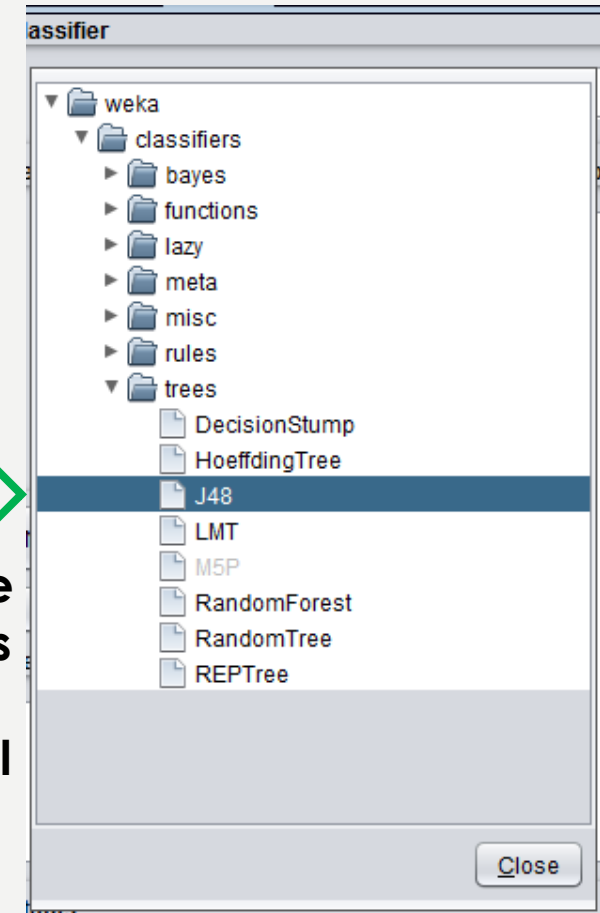
Clic en choose



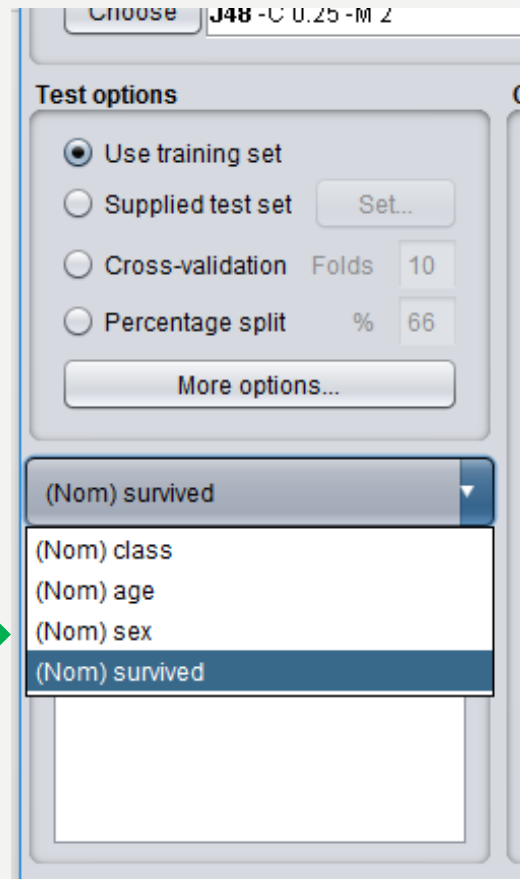
Seleccionamos



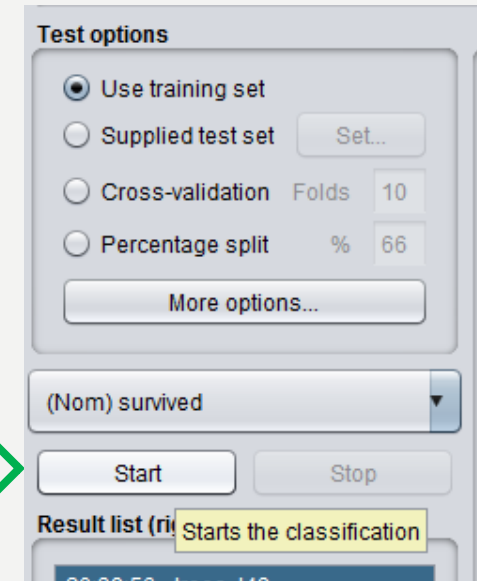
Selección de j48 el cual es el algoritmo para crear el árbol.



Seleccionamos el atributo por el cual realizaremos nuestra búsqueda, en este caso seleccionamos survived.



Clic para comenzar el Análisis.



Se refleja el Análisis.

```
Classifier output

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    relation
Instances:   2201
Attributes:  4
              class
              age
              sex
              survived
Test mode:   evaluate on training data

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

J48 pruned tree
-----

sex = male
|  class = 1st
```

← Total de registro.

← Atributos.

```
Classifier output

-----

sex = male
|  class = 1st
|  |  age = adult: no (175.0/57.0)
|  |  age = child: yes (5.0)
|  class = 2nd
|  |  age = adult: no (168.0/14.0)
|  |  age = child: yes (11.0)
|  class = 3rd no (510.0/88.0)
|  class = crew: no (862.0/192.0)
sex = female
|  class = 1st yes (145.0/4.0)
|  class = 2nd yes (106.0/13.0)
|  class = 3rd no (196.0/90.0)
|  class = crew: yes (23.0/3.0)

Number of Leaves :    10
```

Analisis:sexo Masculino adulto de **primera** clase No sobrevivieron 57 de un total 175 y niño 5.

```
Classifier output
-----
sex = male
| class = 1st
| | age = adult: no (175.0/57.0)
| | age = child: yes (5.0)
| class = 2nd
| | age = adult: no (168.0/14.0)
| | age = child: yes (11.0)
| class = 3rd: no (510.0/88.0)
| class = crew: no (862.0/192.0)
sex = female
| class = 1st: yes (145.0/4.0)
| class = 2nd: yes (106.0/13.0)
| class = 3rd: no (196.0/90.0)
| class = crew: yes (23.0/3.0)

Number of Leaves : 10
```

Analisis: sexo Masculino adulto de **Segunda** clase No sobrevivieron 14 de un total 168 y niño sobrevivieron 11.

```
Classifier output
-----
sex = male
| class = 1st
| | age = adult: no (175.0/57.0)
| | age = child: yes (5.0)
| class = 2nd
| | age = adult: no (168.0/14.0)
| | age = child: yes (11.0)
| class = 3rd: no (510.0/88.0)
| class = crew: no (862.0/192.0)
sex = female
| class = 1st: yes (145.0/4.0)
| class = 2nd: yes (106.0/13.0)
| class = 3rd: no (196.0/90.0)
| class = crew: yes (23.0/3.0)

Number of Leaves : 10
```

Analisis: sexo Masculino adulto de **Tercera** clase No sobrevivieron 88 de un total 510 y tripulación 192 de 862

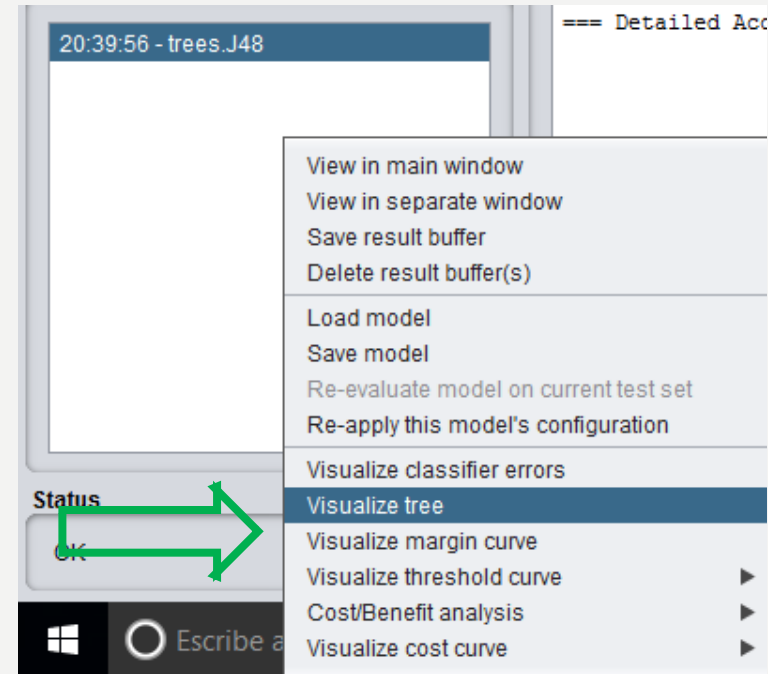
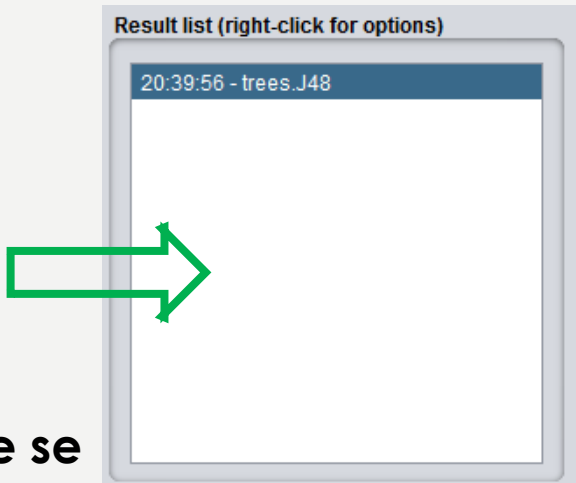
```
Classifier output
-----
sex = male
| class = 1st
| | age = adult: no (175.0/57.0)
| | age = child: yes (5.0)
| class = 2nd
| | age = adult: no (168.0/14.0)
| | age = child: yes (11.0)
| class = 3rd: no (510.0/88.0)
| class = crew: no (862.0/192.0)
sex = female
| class = 1st: yes (145.0/4.0)
| class = 2nd: yes (106.0/13.0)
| class = 3rd: no (196.0/90.0)
| class = crew: yes (23.0/3.0)

Number of Leaves : 10
```

Análisis: sexo femenino de **primera** clase sobrevivieron 4 de 145 de **segunda** clase 13 de 106 de **tercera** clase no sobrevivieron 90 de 196 y **tripulación** sobrevivieron 3 de 23.

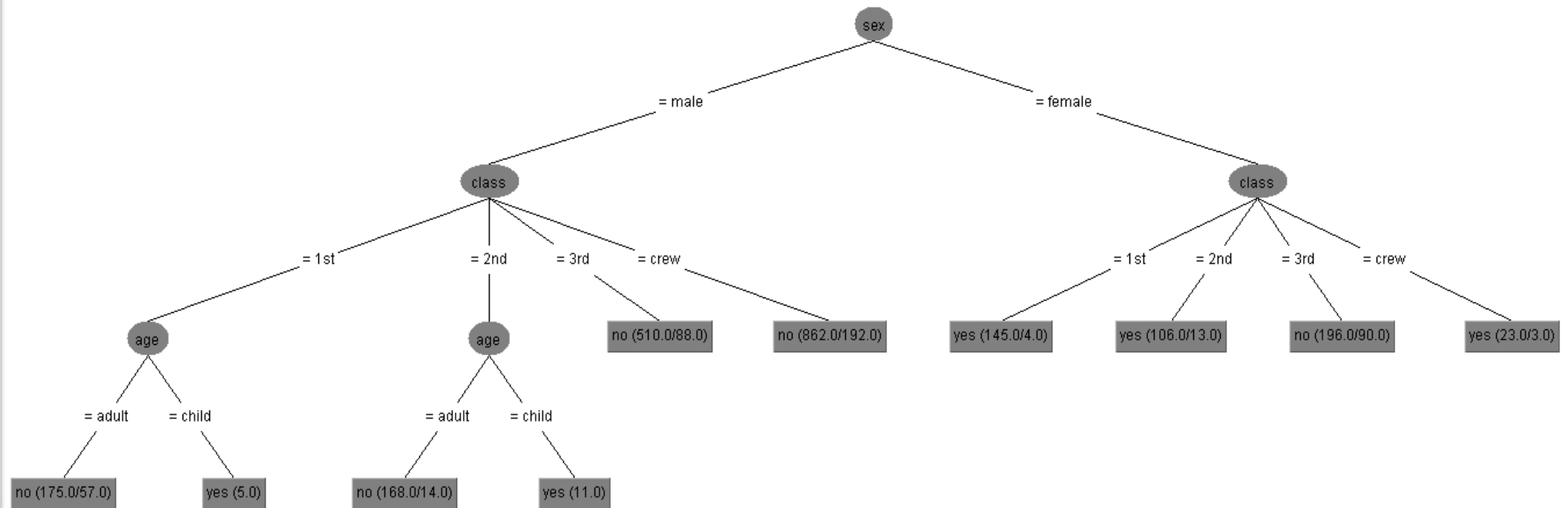
MOSTRAR ÁRBOL DE DECISIÓN

Clic
derecho
para que se
visualice
unas
opciones.



Clic en
visualize
tree

ÁRBOL DE DECISIÓN



BIBLIOGRAFÍA.

- <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/microsoft-decision-trees-algorithm>
- www.se.cuhk.edu.hk/~hcheng/WEKA.ppt