



Programa de Formación: Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos

Octubre 2018

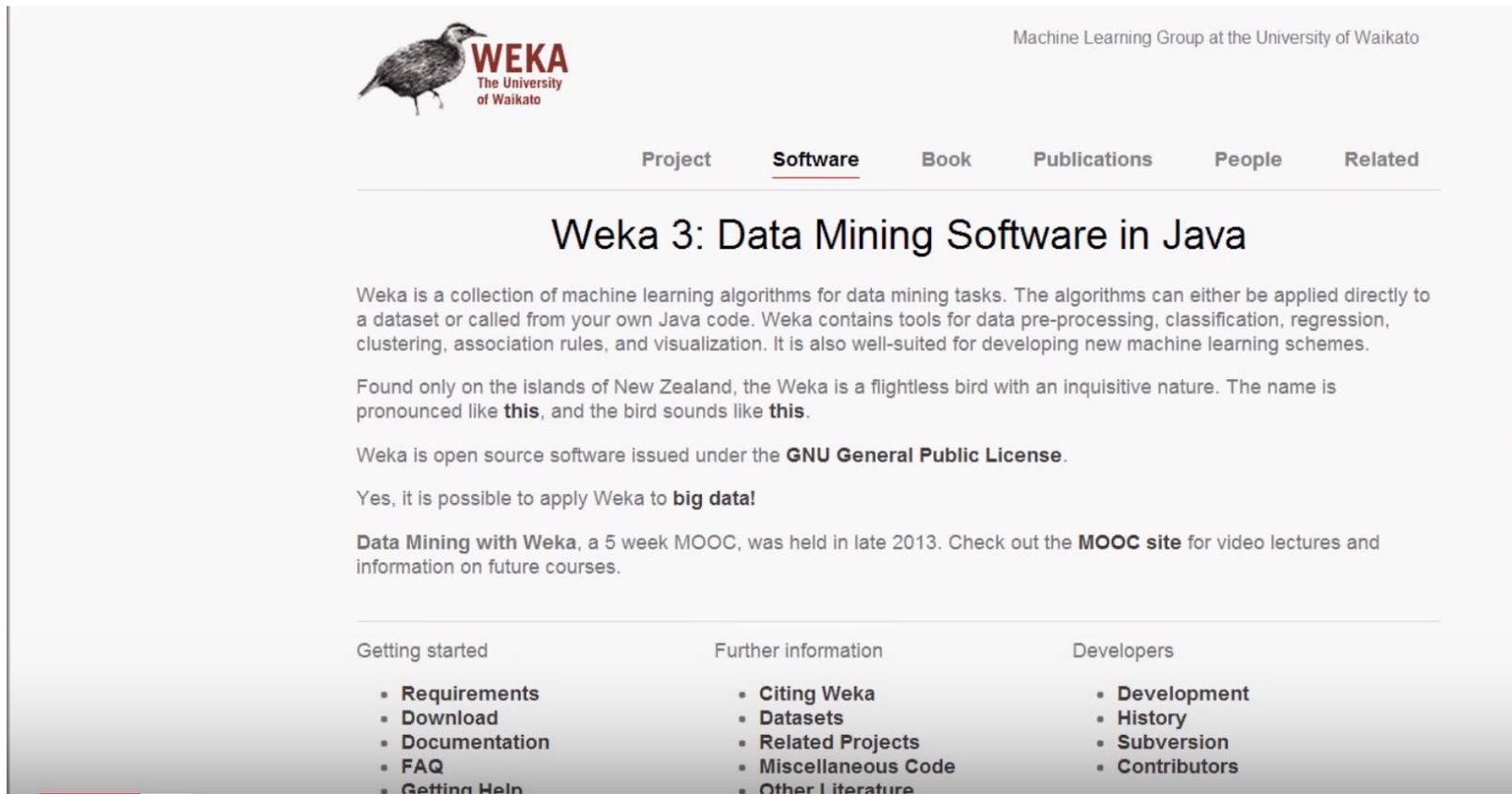
Sesión: Predicción de la rotura de stock (WEKA)
(Parte I: introducción a WEKA)

Profesores:
Alex Rabasa
Carolina Belso

FUNCIONALIDADES DE WEKA y OPCIONES BÁSICAS

- Descarga e instalación de WEKA
- Formato .arff
- Leyendo datos y descriptivas muy básicas
- Datos discretos vs Datos continuos
- Cómo se usan e interpretan diferentes modelos

Descarga e instalación de WEKA



The screenshot shows the WEKA website homepage. At the top left is the WEKA logo featuring a kiwi bird and the text 'WEKA The University of Waikato'. At the top right is the text 'Machine Learning Group at the University of Waikato'. Below the logo is a navigation menu with links for 'Project', 'Software' (which is underlined), 'Book', 'Publications', 'People', and 'Related'. The main heading is 'Weka 3: Data Mining Software in Java'. Below this is a paragraph describing WEKA as a collection of machine learning algorithms for data mining tasks. Another paragraph explains the origin of the name 'Weka' from a flightless bird in New Zealand. A third paragraph states that WEKA is open source software under the GNU General Public License. A fourth paragraph says 'Yes, it is possible to apply Weka to big data!'. A fifth paragraph mentions a MOOC course held in late 2013. At the bottom, there are three columns of links: 'Getting started' (Requirements, Download, Documentation, FAQ, Getting Help), 'Further information' (Citing Weka, Datasets, Related Projects, Miscellaneous Code, Other Literature), and 'Developers' (Development, History, Subversion, Contributors).

Machine Learning Group at the University of Waikato

[Project](#) [Software](#) [Book](#) [Publications](#) [People](#) [Related](#)

Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

Yes, it is possible to apply Weka to **big data!**

Data Mining with Weka, a 5 week MOOC, was held in late 2013. Check out the **MOOC site** for video lectures and information on future courses.

Getting started	Further information	Developers
<ul style="list-style-type: none">• Requirements• Download• Documentation• FAQ• Getting Help	<ul style="list-style-type: none">• Citing Weka• Datasets• Related Projects• Miscellaneous Code• Other Literature	<ul style="list-style-type: none">• Development• History• Subversion• Contributors

Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos

Formato .arff

```
@relation credit-rating
```

```
@attribute A1          {b,a}
@attribute A2          REAL
@attribute A3          REAL
@attribute A4          {u, y, l, t}
@attribute A5          {g, p, gg}
@attribute A6          {c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff}
@attribute A7          {v, h, bb, j, n, z, dd, ff, o}
@attribute A8          REAL
@attribute A9          {t,f}
@attribute A10         {t,f}
@attribute A11         REAL
@attribute A12         {t,f}
@attribute A13         {g, p, s}
@attribute A14         REAL
@attribute A15         REAL
@attribute class       {+, -}
```

```
@data
```

```
%
```

```
% 690 instances
```

```
%
```

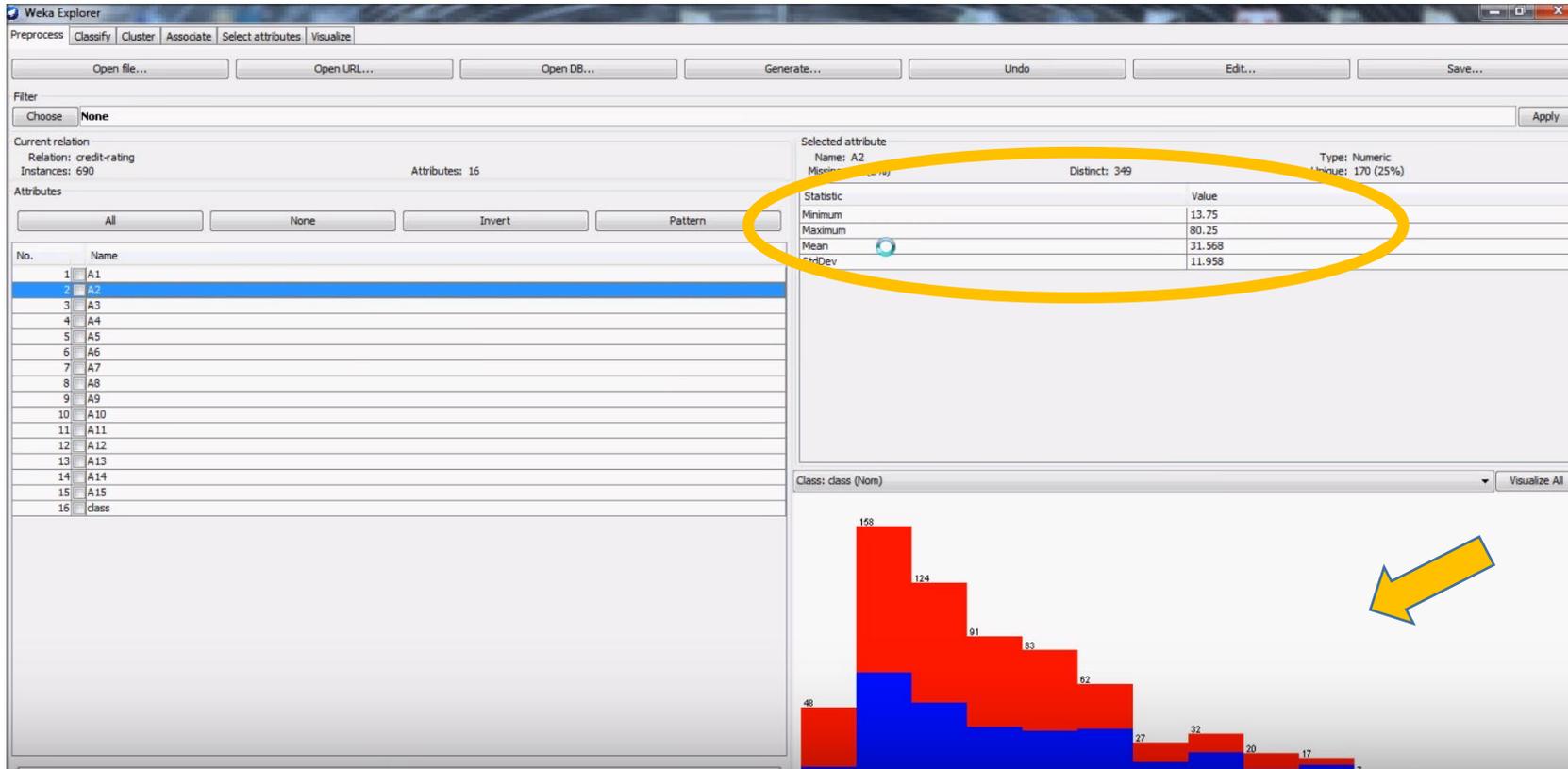
```
b,30.83,0,u,g,w,v,1.25,t,t,01,f,g,00202,0,+
a,58.67,4.46,u,g,q,h,3.04,t,t,06,f,g,00043,560,+
a,24.50,0.5,u,g,q,h,1.5,t,f,0,f,g,00280,824,+
b,27.83,1.54,u,g,w,v,3.75,t,t,05,t,g,00100,3,+
b,20.17,5.625,u,g,w,v,1.71,t,f,0,f,s,00120,0,+
b,32.08,4,u,g,m,v,2.5,t,f,0,t,g,00360,0,+
```

Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos

<https://www.youtube.com/watch?v=VZv4HzNiWj0>

(ejemplo paso a paso)

Leyendo datos y descriptivas muy básicas



Datos discretos vs Datos continuos

Selected attribute
Name: A1
Missing: 12 (2%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

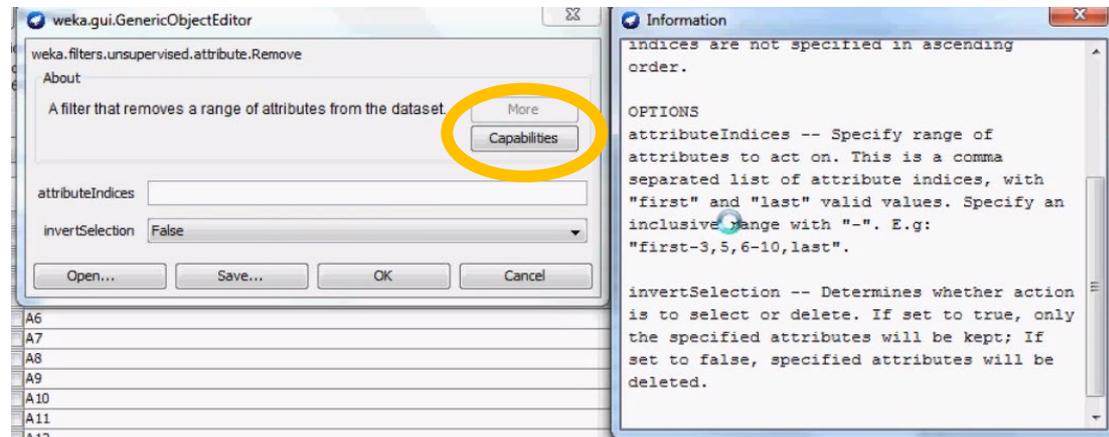
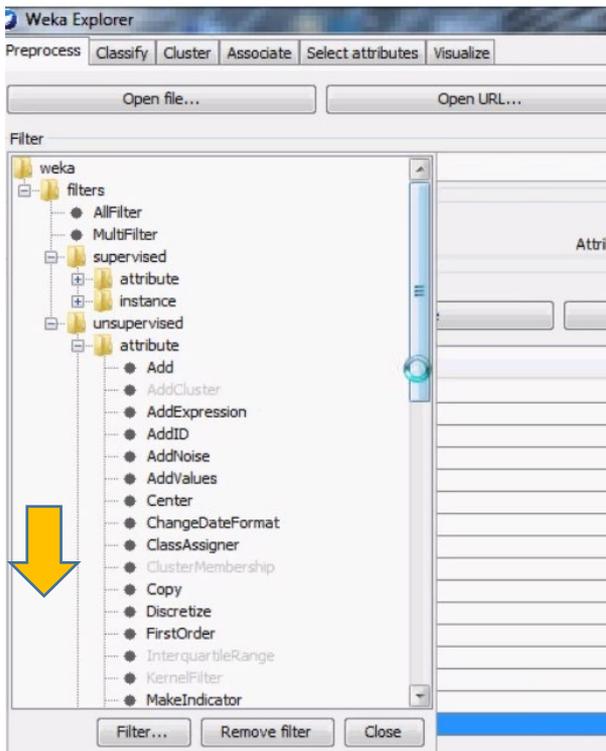
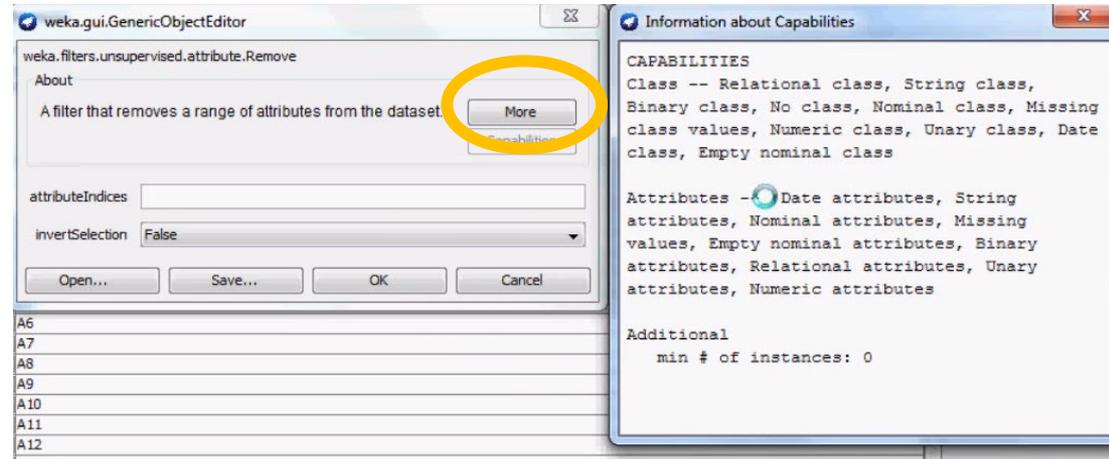
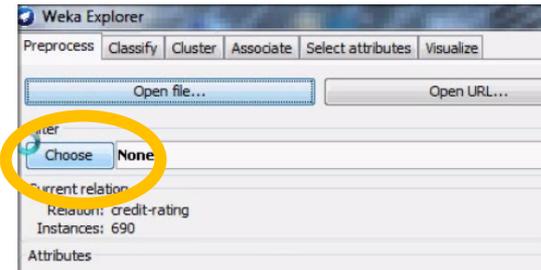
No.	Label	Count
1	b	468
2	a	210

Selected attribute
Name: A2
Missing: 12 (2%)
Distinct: 349
Type: Numeric
Unique: 170 (25%)

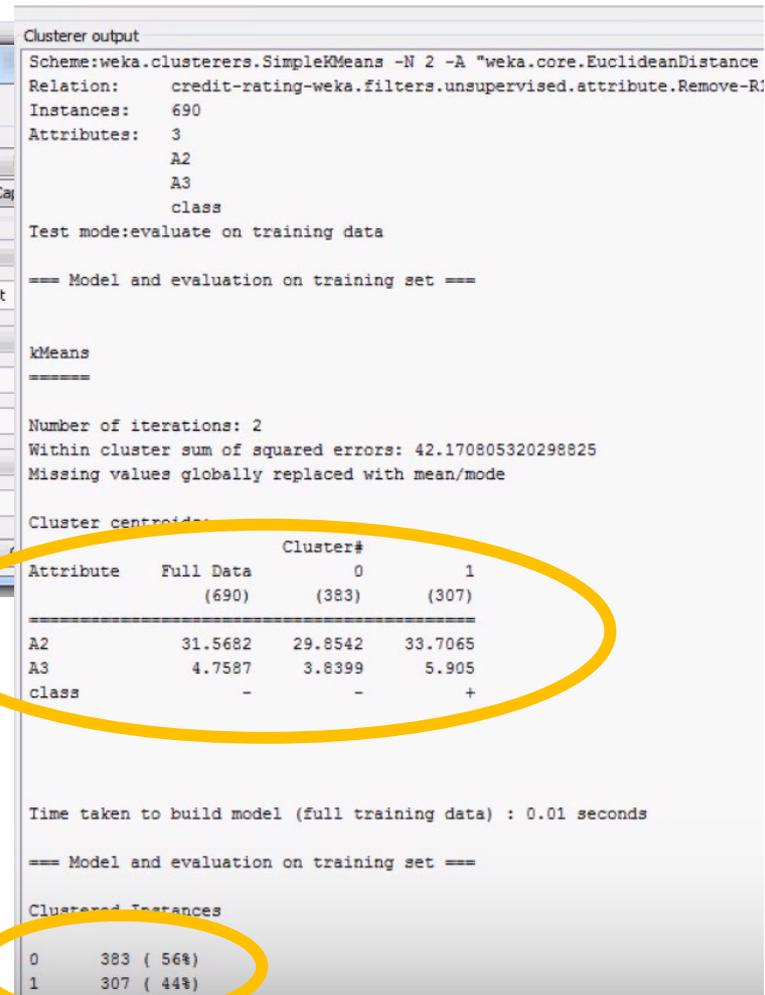
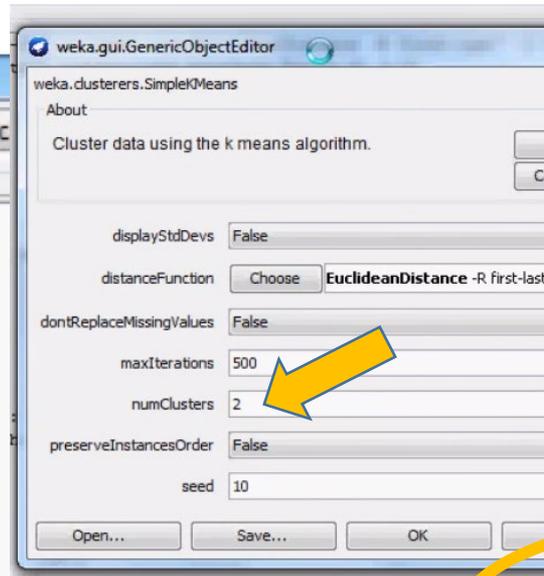
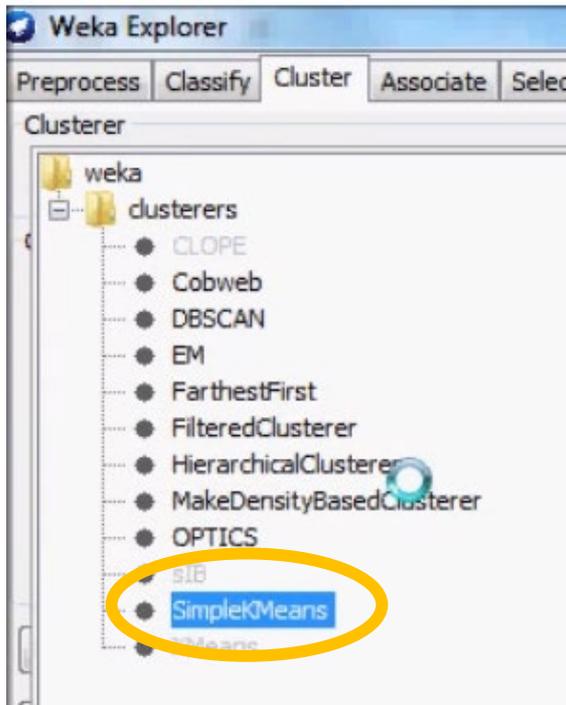
Statistic	Value
Minimum	13.75
Maximum	80.25
Mean	31.568
StdDev	11.958

WEKA permitirá usar unos métodos u otros, según la naturaleza de los datos.

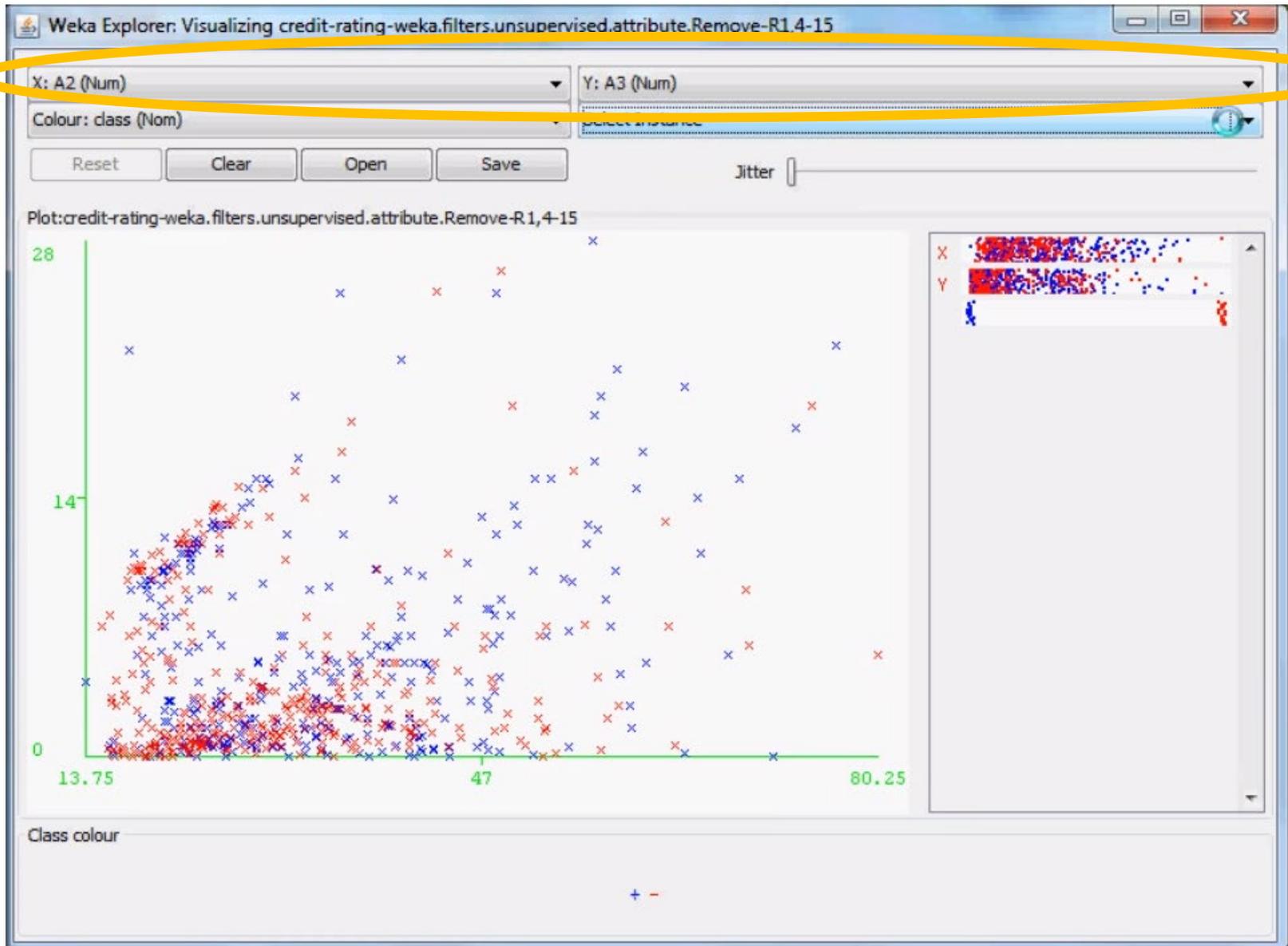
Cómo se usan e interpretan diferentes modelos



CLUSTERING o SEGMENTACIÓN (con Simple Kmeans)

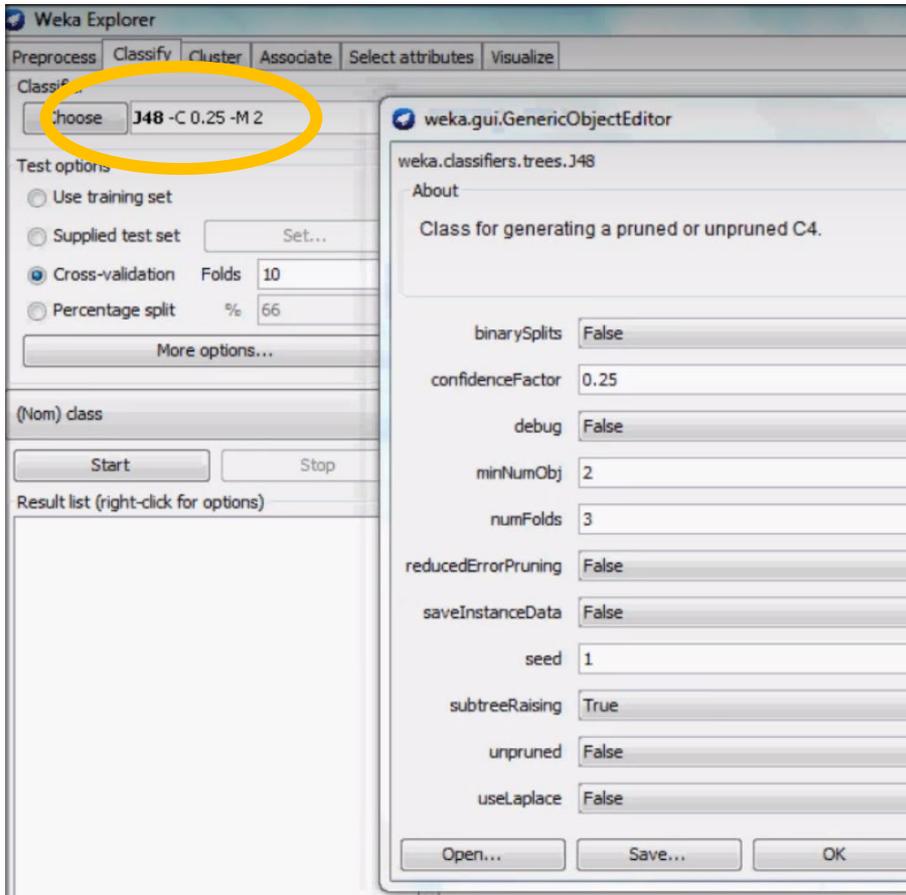


Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos



Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos

CLASIFICACIÓN (con J48)



Classifier output

```
| | A7 = j: + (1.24/0.08)
| | A7 = n: + (1.24/0.08)
| | A7 = z: - (0.0)
| | A7 = dd: - (0.0)
| | A7 = ff: - (4.96/0.64)
| | A7 = o: - (0.0)
| A3 > 0.165: - (298.0/12.0)

Number of Leaves : 30
Size of the tree : 42

Time taken to build model: 0.09 seconds

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

Correctly Classified Instances 594 86.087 %
Incorrectly Classified Instances 96 13.913 %
Kappa statistic 0.718
Mean absolute error 0.1924
Root mean squared error 0.3313
Relative absolute error 38.9417 %
Root relative squared error 66.6637 %
Total Number of Instances 690

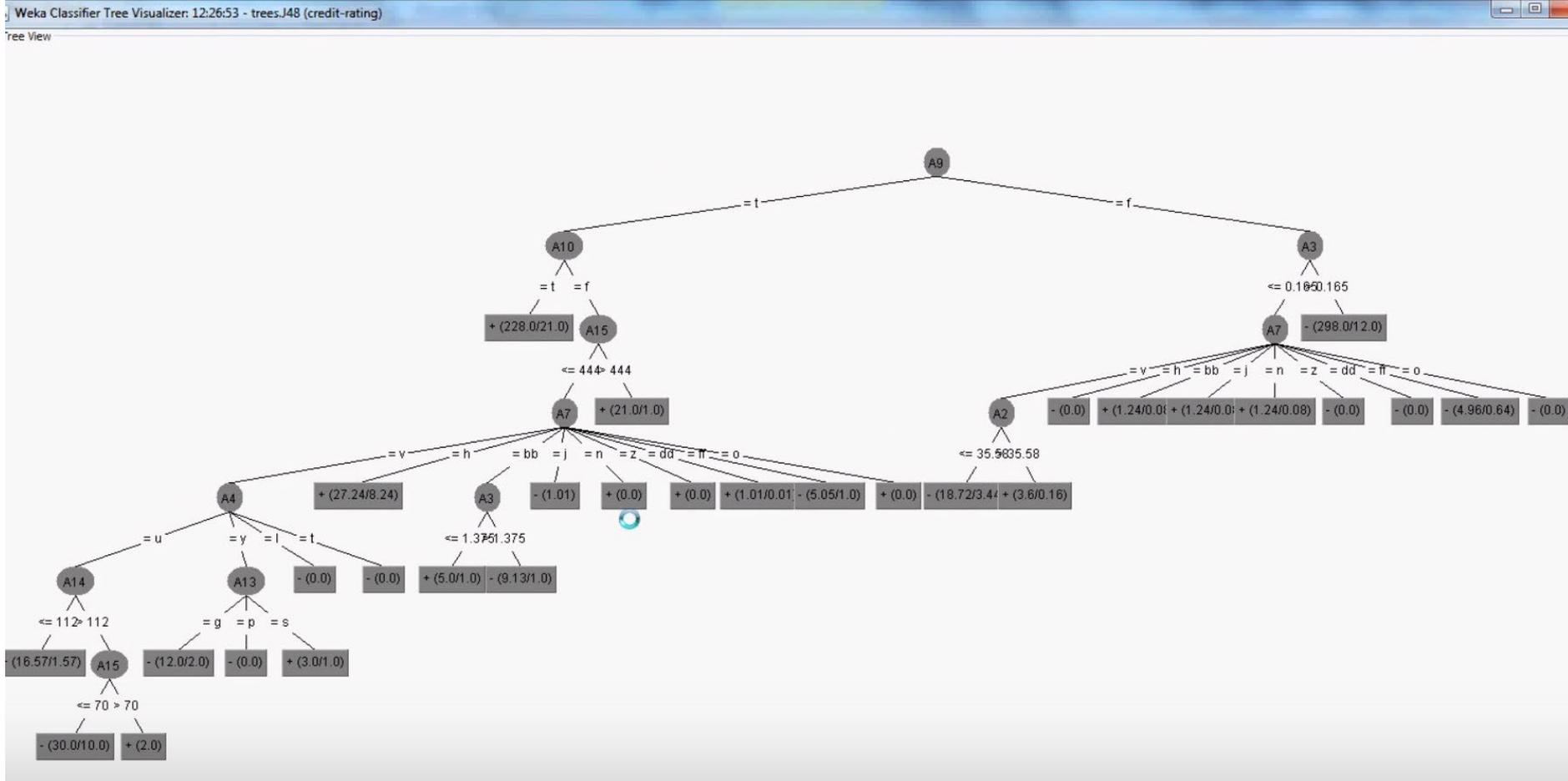
=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.837   0.12    0.848    0.837   0.843    0.887    +
          0.88    0.163   0.871    0.88    0.875    0.887    -
Weighted Avg.  0.861   0.144   0.861    0.861   0.861    0.887

=== Confusion Matrix ===

 a  b  <-- classified as
257 50 | a = +
 46 337 | b = -
```

Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos



Aplicación del Big Data en el sector del calzado para la mejora de la competitividad y de los procesos productivos



GRACIAS