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

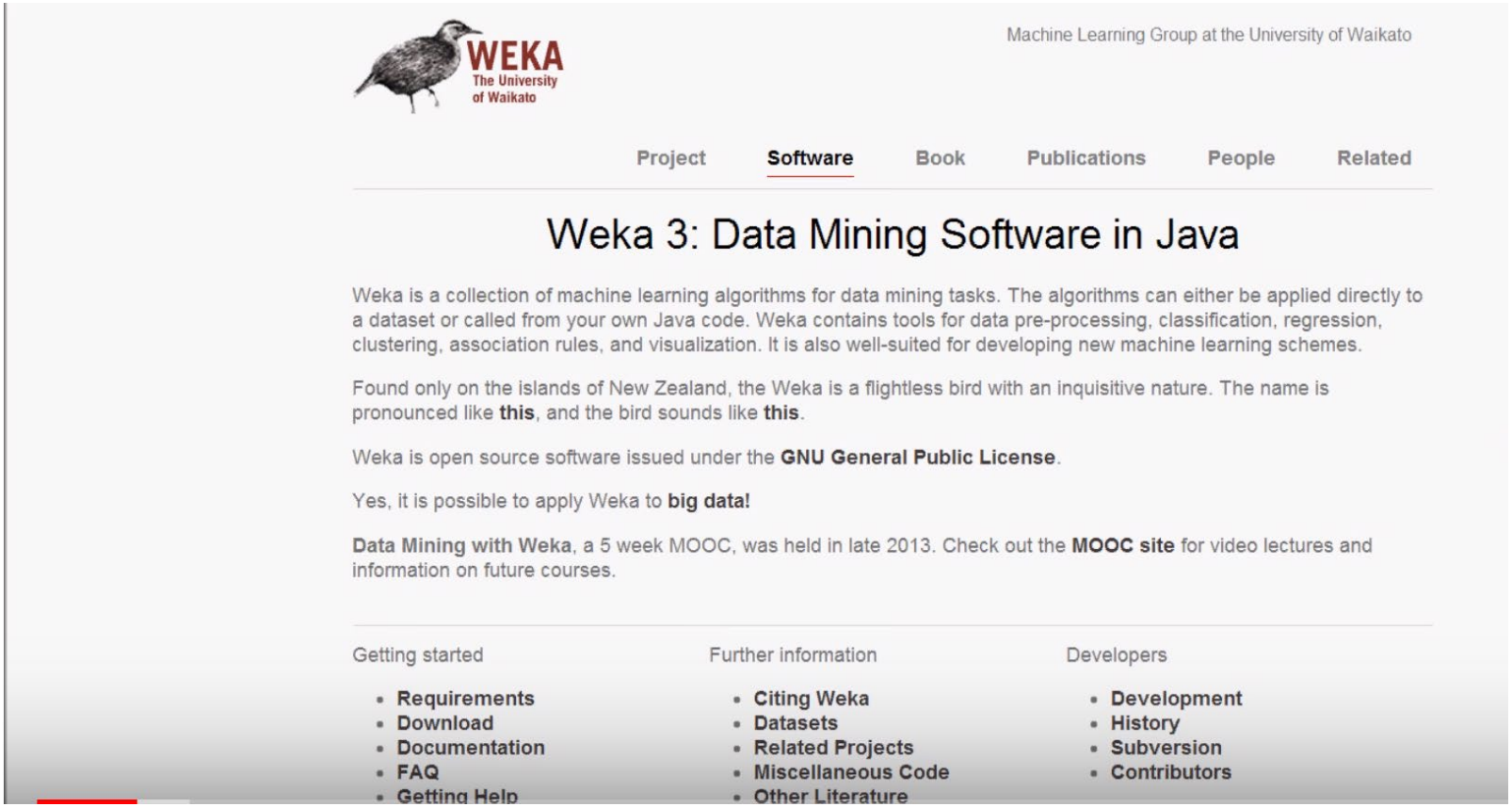
Sesión: Predicción de la rotura de stock (WEKA)
(Parte I: introducción a WEKA)

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

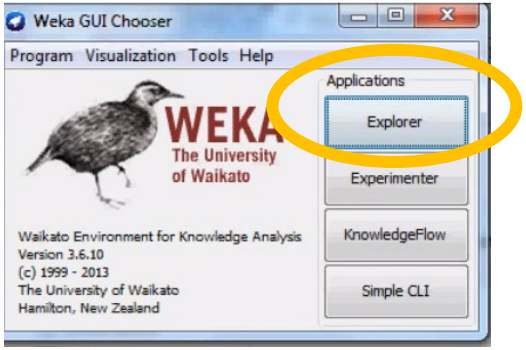


Formato .arff

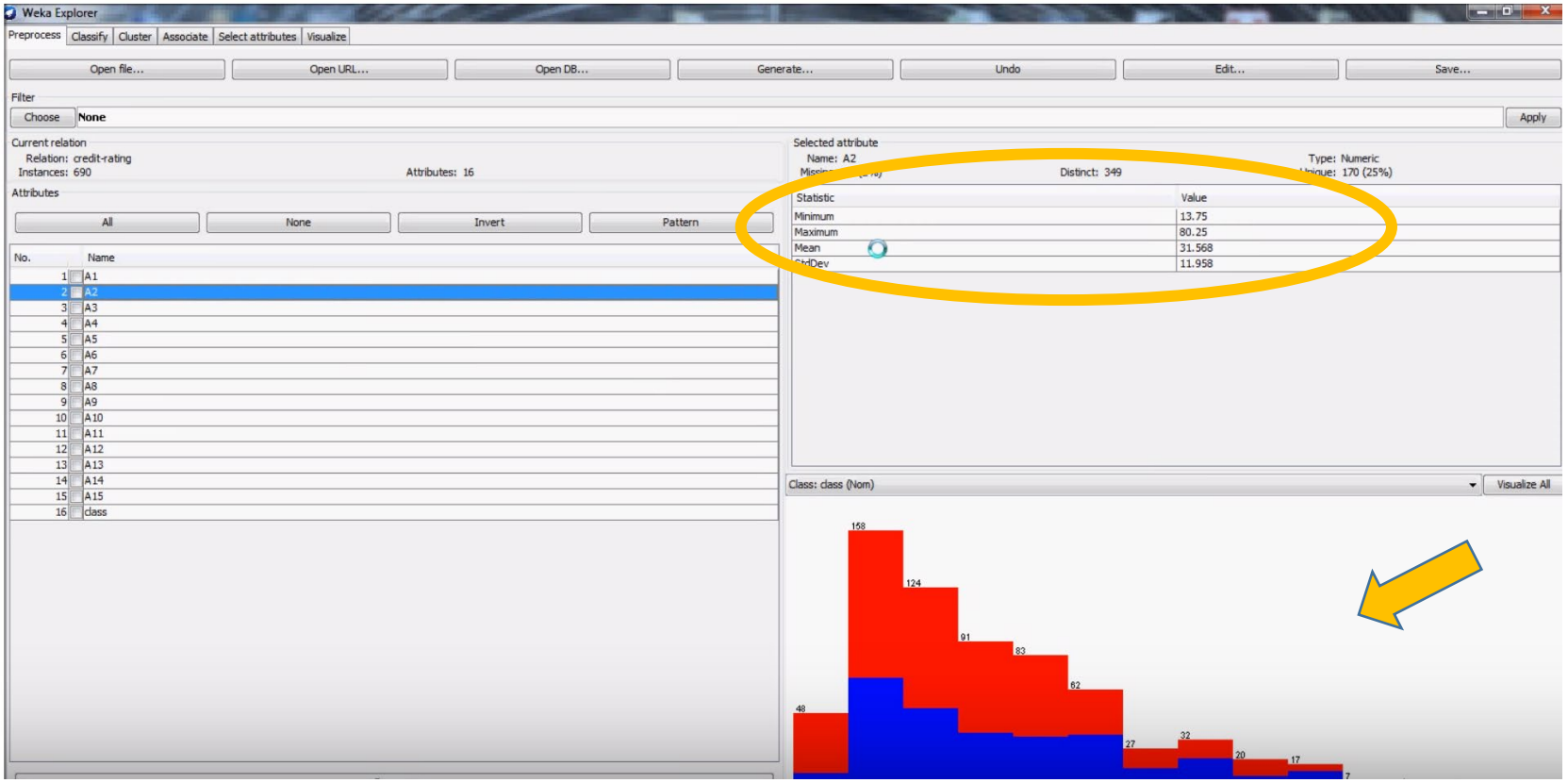
```
@relation credit-rating

@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,+
```

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%)

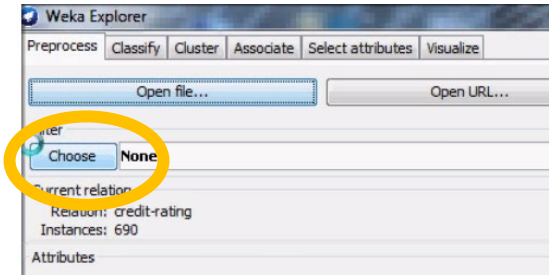
No.	Label	Count
1	b	468
2	a	210

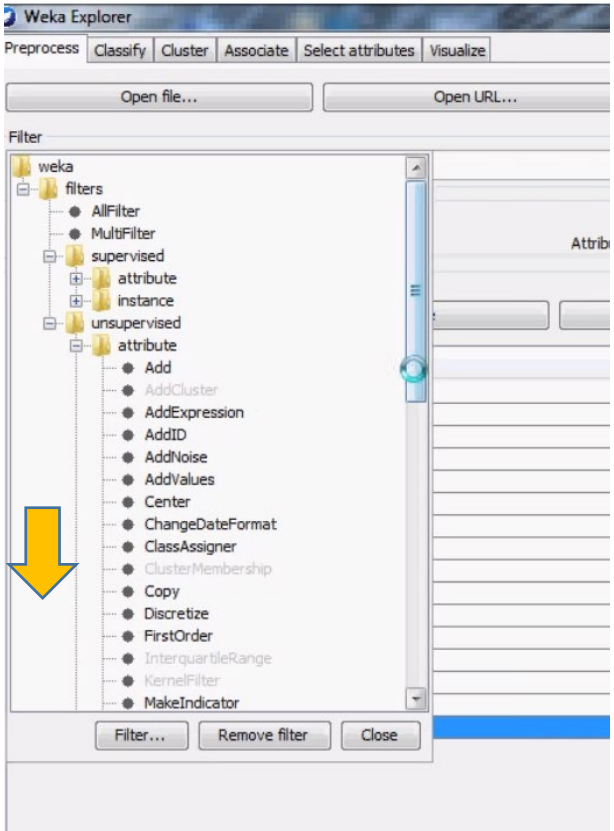
Selected attribute
Name: A2
Missing: 12 (2%)
Distinct: 349
Type: Numeric
Unique: 170 (25%)

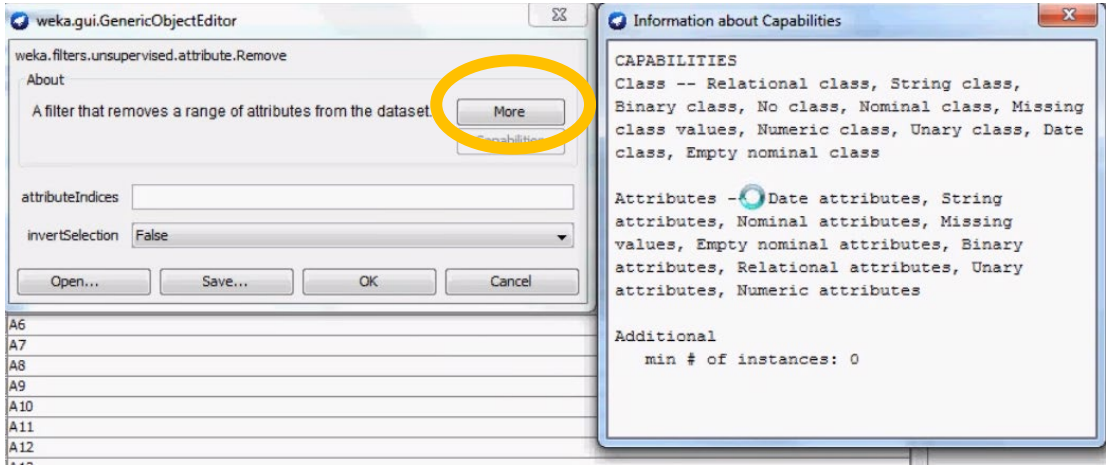
Statistic	Value
Minimum	13.75
Maximum	80.25
Mean	31.568
StdDev	11.958

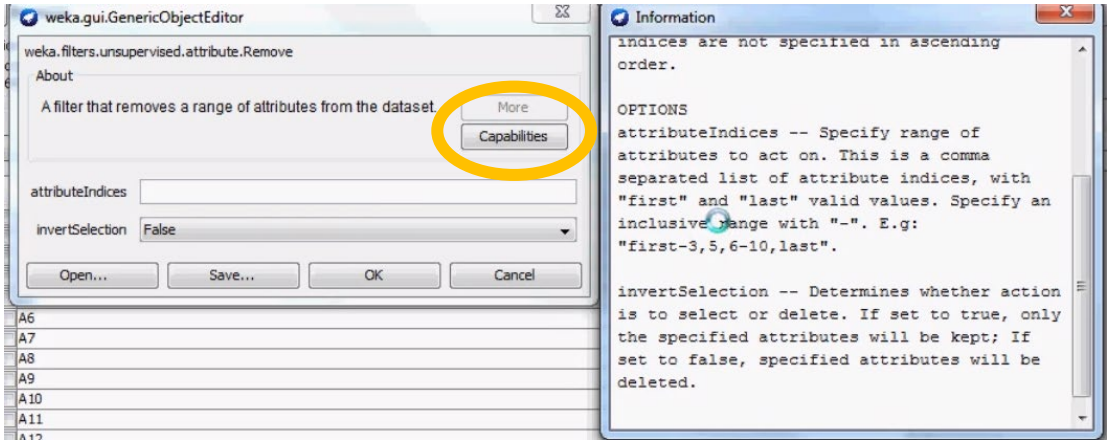
WEKA permitirá usar unos métodos u otros, según la naturaleza de los datos.

Cómo se usan e interpretan diferentes modelos

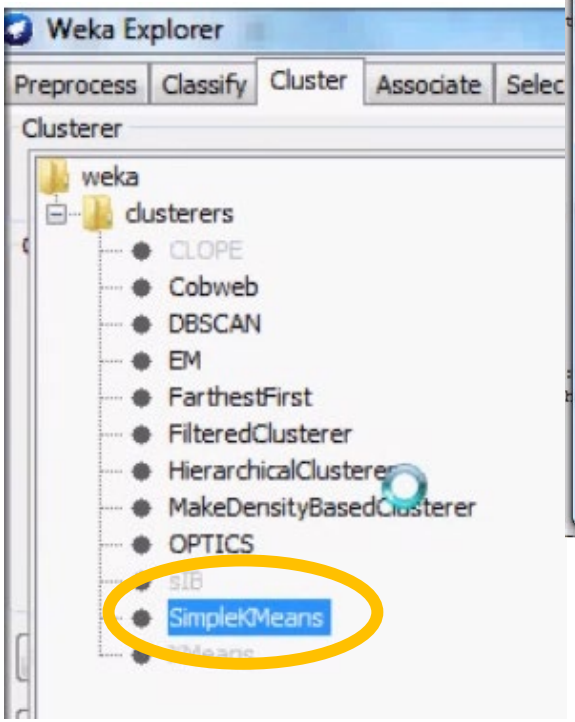


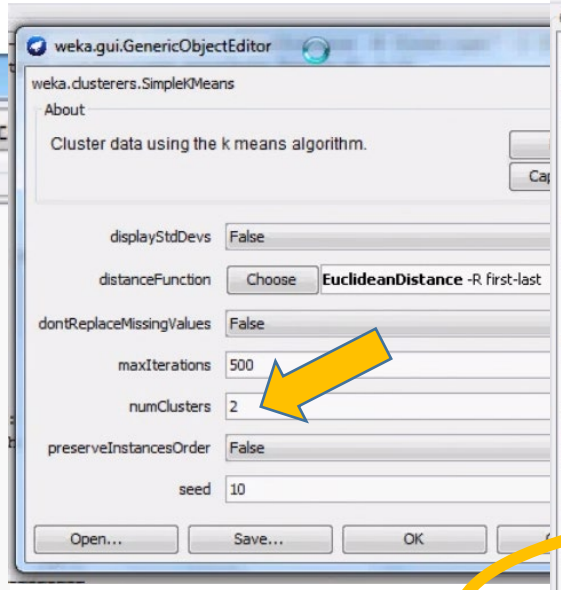


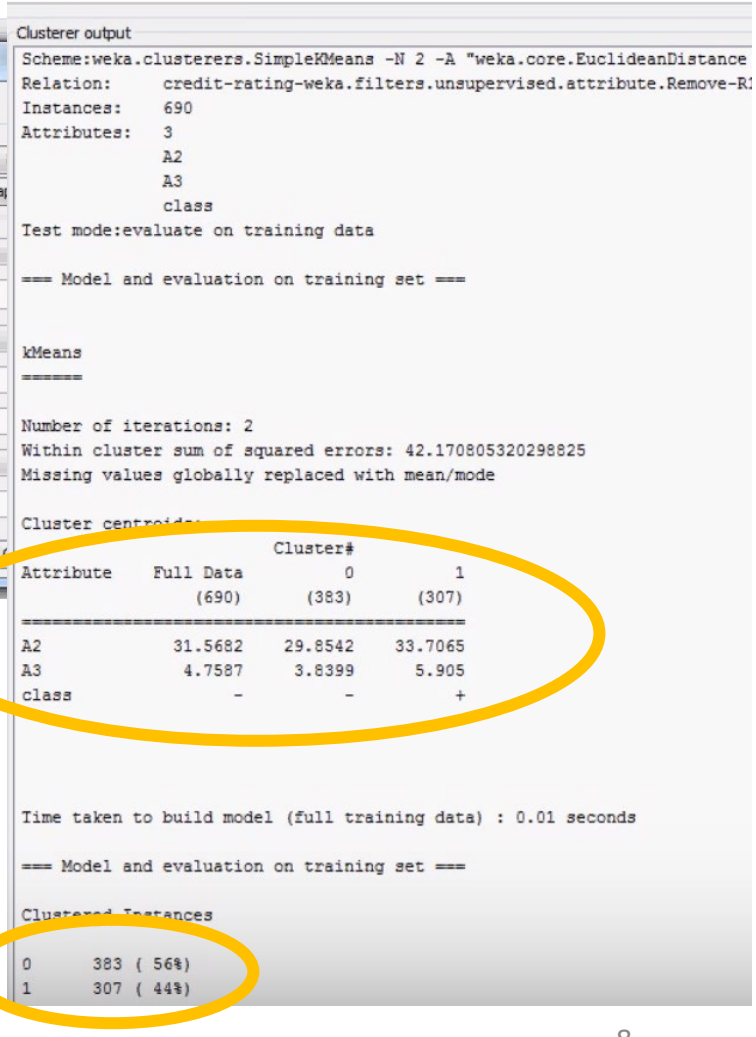




CLUSTERING o SEGMENTACIÓN (con Simple Kmeans)

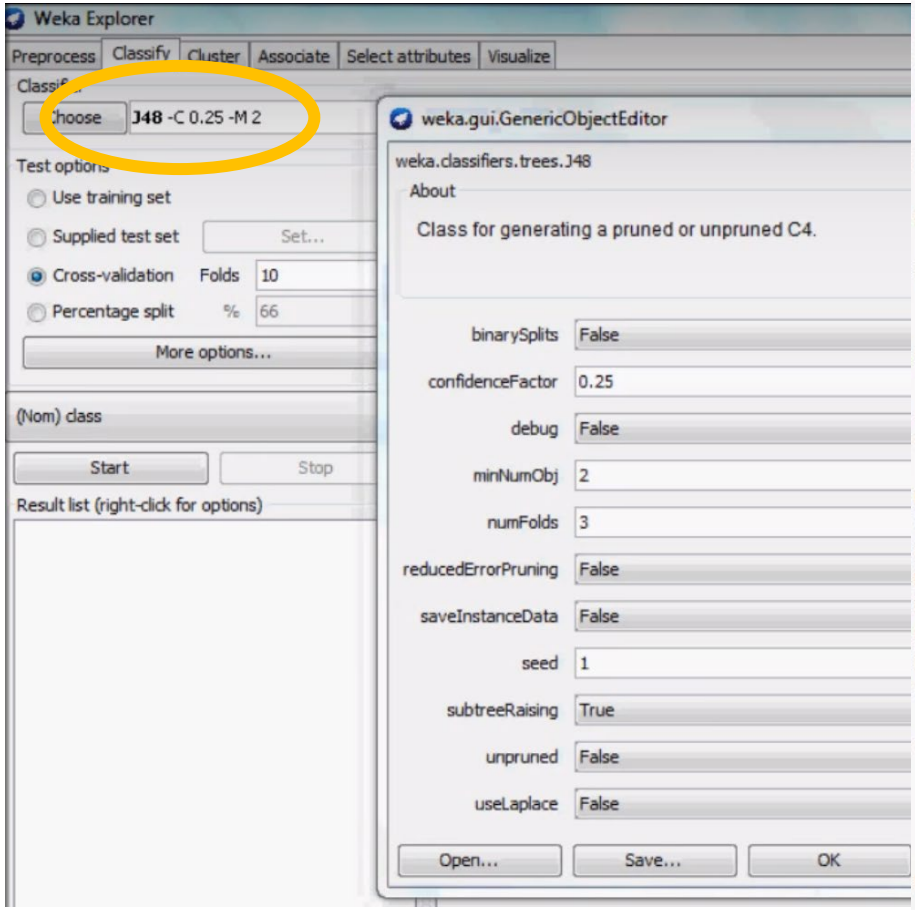






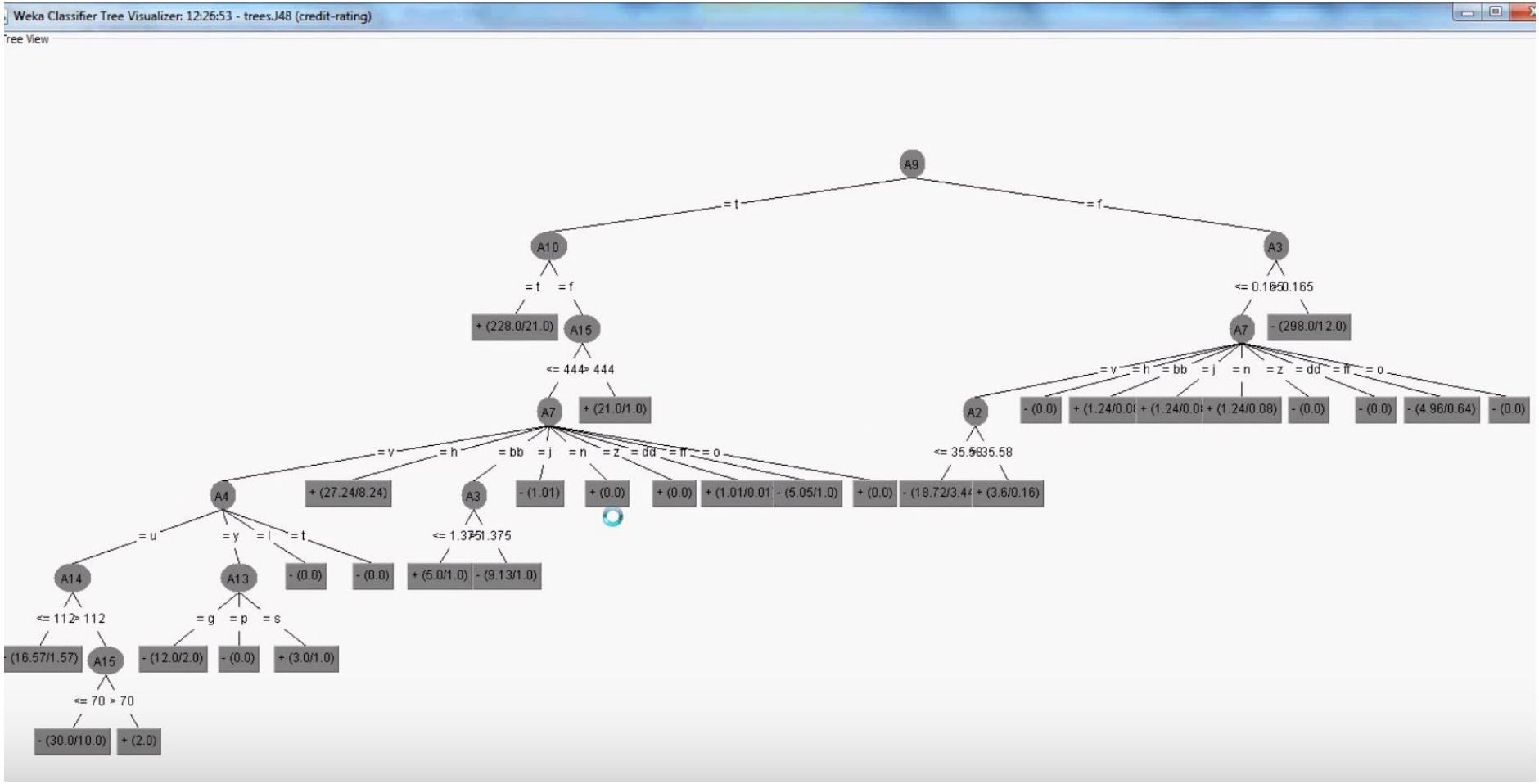

Weka Explorer: Visualizing credit-rating-weka.filters.unsupervised.attribute.Remove-R1,4-15
X: A2 (Num)
Y: A3 (Num)
Colour: class (Nom)
Reset
Clear
Open
Save
Jitter
Plot:credit-rating-weka.filters.unsupervised.attribute.Remove-R1,4-15
28
14
0
13.75
47
80.25
X
Y
Class colour
+ -

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 = -
```


Weka Classifier Tree Visualizer: 12:26:53 - trees.J48 (credit-rating)
Tree View
A9
= t
= f
A10
A3
= t
= f
<= 0.165
+ (228.0/21.0)
A15
A7
- (298.0/12.0)
<= 444
> 444
A7
+ (21.0/1.0)
A2
- (0.0)
= v
= h
= bb
= j
= n
= z
= dd
= o
<= 35.58
A4
+ (27.24/8.24)
A3
- (1.01)
+ (0.0)
+ (0.0)
- (5.05/1.0)
+ (0.0)
+ (3.6/0.16)
= u
= y
= l
= t
<= 1.375
A14
A13
- (0.0)
- (0.0)
+ (5.0/1.0)
- (9.13/1.0)
<= 112
> 112
= g
= p
= s
- (16.57/1.57)
A15
- (12.0/2.0)
- (0.0)
+ (3.0/1.0)
<= 70
> 70
- (30.0/10.0)
+ (2.0)

GRACIAS