



Alumno:

Juan Cañar U.

Materia:

SE

Ciclo:

9no

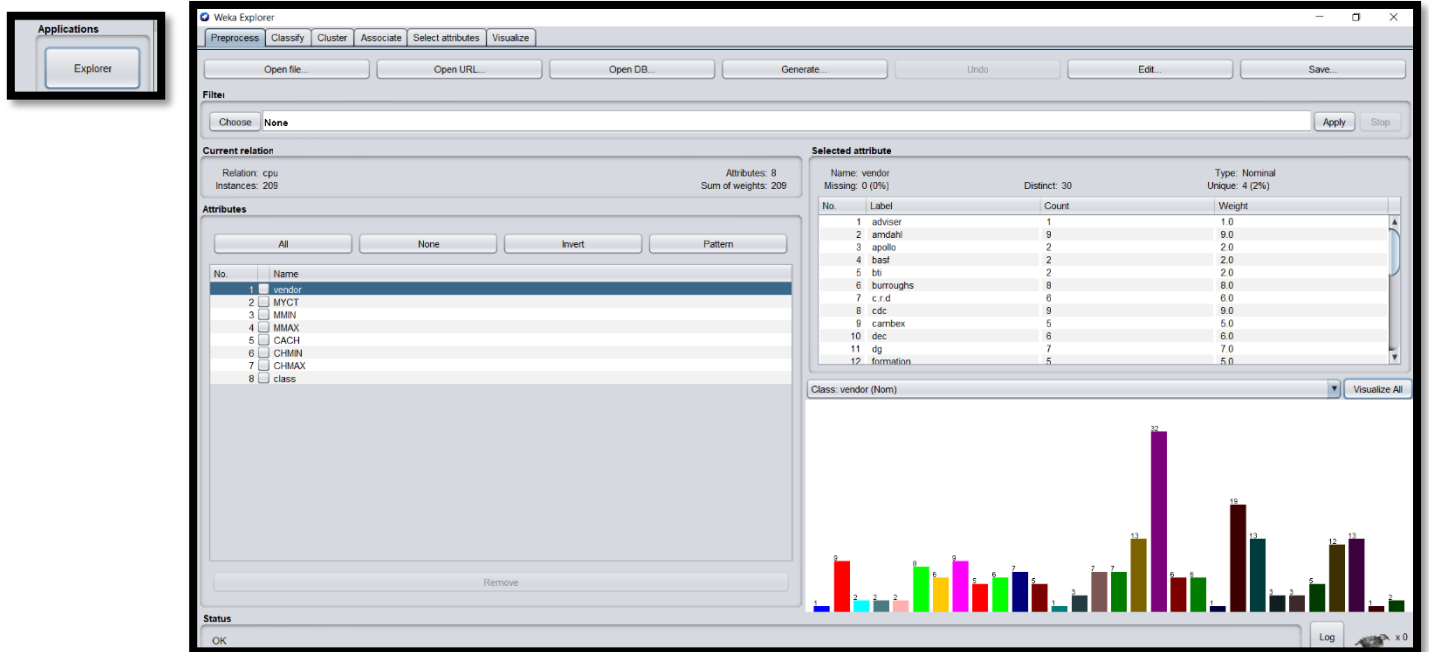
Fecha:

24/07/2020

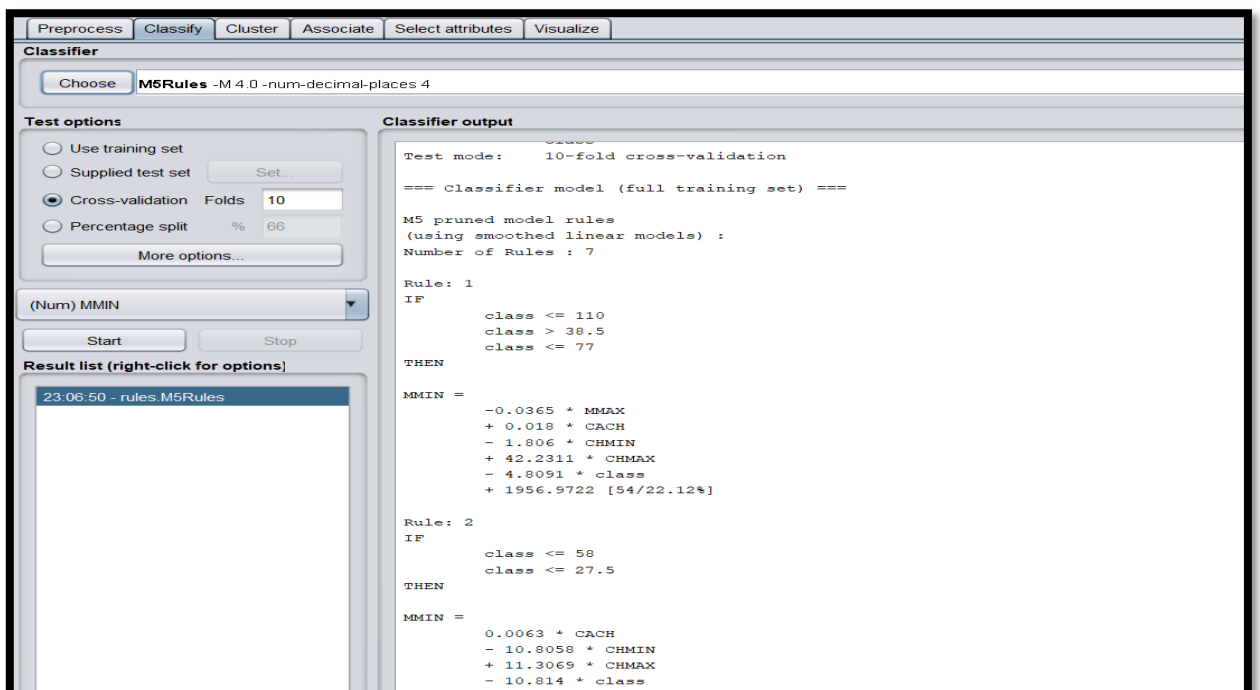
- **Que es weka?**

Entorno para análisis del conocimiento de la Universidad de Waikato, es una plataforma de software para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato.

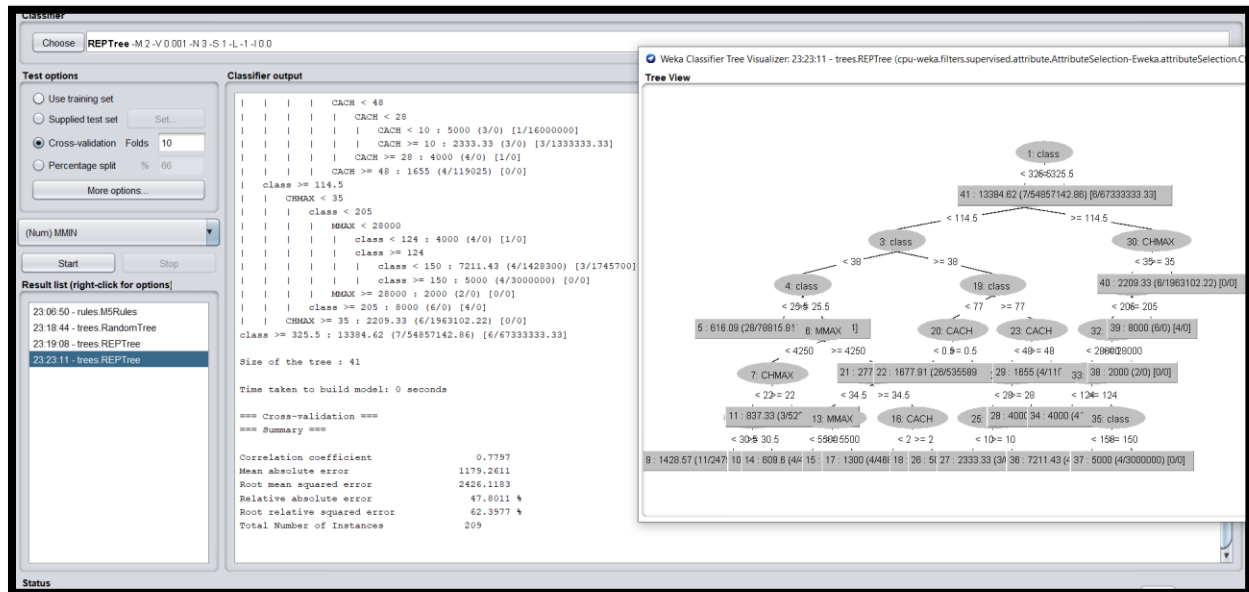
1. Realizamos la carga de datos, en weka, para esto se eligió el dataset, “cpu.with.vendor.arff”.



2. Analizamos los datos mediante reglas, existen distintas opciones para poder realizar, la clasificación.



- Se puede generar un árbol clasificatorio para poder entender como están los datos, en este caso el árbol es generado a partir de valor máximos del grupo de mínimos.
 - Como se puede ver en el árbol hay una clasificatoria donde nos dice los porcentajes de datos analizados.



- Se puede ver en el grafico a continuación como clasifica lo que vendió en porcentajes, y las marcas que vendió.

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) vendor

Start Stop

Result list (right-click for options)

23:54:44 - trees.J48

23:55:47 - trees.J48

Classifier output

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,010	0,333	1,000	0,500	0,500	0,575	0,995	0,333	adviser
0,889	0,005	0,889	0,889	0,889	0,889	0,884	0,996	0,857	amdahl
0,500	0,005	0,500	0,500	0,500	0,500	0,495	0,981	0,313	apollo
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	basf
0,500	0,005	0,500	0,500	0,500	0,500	0,495	0,995	0,500	bti
1,000	0,010	0,800	1,000	0,889	0,890	0,890	0,996	0,843	burroughs
0,833	0,005	0,833	0,833	0,833	0,828	0,828	0,996	0,806	c.r.d
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	cdc
1,000	0,005	0,833	1,000	0,909	0,911	0,908	0,998	0,833	cambex
0,500	0,010	0,600	0,500	0,545	0,536	0,536	0,973	0,474	dec
1,000	0,005	0,875	1,000	0,933	0,933	0,933	0,999	0,964	dg
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	formation
0,000	0,000	?	0,000	?	?	?	0,995	0,333	four-phase
1,000	0,005	0,750	1,000	0,857	0,864	0,864	0,998	0,750	gould
0,714	0,000	1,000	0,714	0,833	0,841	0,841	0,996	0,868	hp
1,000	0,005	0,875	1,000	0,933	0,933	0,933	0,999	0,946	harris
0,923	0,000	1,000	0,923	0,960	0,958	0,958	0,999	0,976	honeywell
0,938	0,034	0,833	0,938	0,882	0,862	0,862	0,990	0,924	ibm
0,667	0,005	0,800	0,667	0,727	0,723	0,723	0,985	0,745	ipl
0,833	0,005	0,833	0,833	0,833	0,828	0,828	0,998	0,875	magnuson
0,000	0,000	?	0,000	?	?	?	0,995	0,333	microdata
0,895	0,032	0,739	0,895	0,810	0,793	0,793	0,984	0,793	nas
0,846	0,005	0,917	0,846	0,880	0,873	0,873	0,994	0,893	ncr
0,667	0,000	1,000	0,667	0,800	0,815	0,815	0,999	0,917	nixdorf
0,667	0,005	0,667	0,667	0,667	0,662	0,662	0,996	0,644	perkin-elmer
0,800	0,005	0,800	0,800	0,800	0,795	0,795	0,995	0,740	prime
0,667	0,005	0,889	0,667	0,762	0,758	0,758	0,982	0,790	siemens
0,769	0,010	0,833	0,769	0,800	0,788	0,788	0,984	0,771	sperry
0,000	0,000	?	0,000	?	?	?	0,995	0,333	status
0,000	0,000	?	0,000	?	?	?	0,981	0,313	wang
Weighted Avg.	0,837	0,011	?	0,837	?	?	0,992	0,833	

Status

- Genera un entrenamiento previo con los datos: Donde nos dice que tiene una predicción de 175 correctos, esto equivale al 83.73%, la clasificación de incorrectos es de 34 que equivale al 16.26%, en el dataset existe un total de 209.

```

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances      175           83.7321 %
Incorrectly Classified Instances    34           16.2679 %
Kappa statistic                    0.8261
Mean absolute error                 0.0156
Root mean squared error             0.0883
Relative absolute error             24.8196 %
Root relative squared error         49.9004 %
Total Number of Instances          209

```

- Aquí nos indica que realizo un total de 8 iteraciones, donde nos clasifica el vendedor, los porcentajes de venta y las marcas que han sido vendidas, este clasifica en dos clústeres.

Clusterer
Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500

Cluster mode
☒ Use training set
☐ Supplied test set Set...
☐ Percentage split % 66
☐ Classes to clusters evaluation (Num) class
☒ Store clusters for visualization
Ignore attributes
Start Stop

Result list (right-click for options)
00:26:08 - SimpleKMeans

Clusterer output

```

KMeans
=====

Number of iterations: 8
Within cluster sum of squared errors: 182.0143438062634

Initial starting points (random):

Cluster 0: ibm,600,768,2000,0,1,1,20
Cluster 1: siemens,59,4000,12000,32,6,12,81

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (209.0)      (155.0)      (54.0)
=====
vendor         ibm          ibm          nas
MYCT           203.823      255.1935     56.3704
MMIN           2867.9809    1671.3032    6302.8889
MMAV           11796.1531   7453.0065    24262.5926
CACH           25.2057      9.2194       71.0926
CHMIN          4.6986       2.8065       10.1296
CHMAX          18.2679      12.7355      34.1481
class          99.3301      47.3871      248.4259

```

```

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      155 ( 74%)
1       54 ( 26%)

```

- Aquí se realiza la selección de atributos, donde nos indica que el 2 y el 6 son los datos que realmente tienen importancia, donde el 2 es el más importante,

The screenshot shows the WEKA Attribute Evaluator window. The 'Select attributes' tab is active. The 'Attribute Evaluator' section has 'Choose' set to 'CfsSubsetEval -P 1 -E 1'. The 'Search Method' section has 'Choose' set to 'BestFirst -D 1 -N 5'. In the 'Attribute Selection Mode' section, 'Use full training set' is selected, with 'Folds' set to 10 and 'Seed' set to 1. The 'Attribute selection output' pane shows the following text:

```

MYCT
MMIN
MMAX
CACH
CHMIN
CHMAX
class
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 27
  Merit of best subset found:    0.617

Attribute Subset Evaluator (supervised, Class (nominal): 1 vendor):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 2,6 : 2
                    MYCT
                    CHMIN

```

The 'Result list (right-click for options)' section shows a single entry: '00:37:04 - BestFirst + CfsSubsetEval'.

- Volvemos a seleccionar el segundo atributo, CHMIN y nos tendría que dar el resultado de vendedor, ya que nuestro interés ha sido analizar las ventas de cada uno y la marca,

The screenshot shows the WEKA Attribute Evaluator window after selecting the second attribute, CHMIN. The 'Attribute selection output' pane shows the following text:

```

Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 41
  Merit of best subset found:    0.66

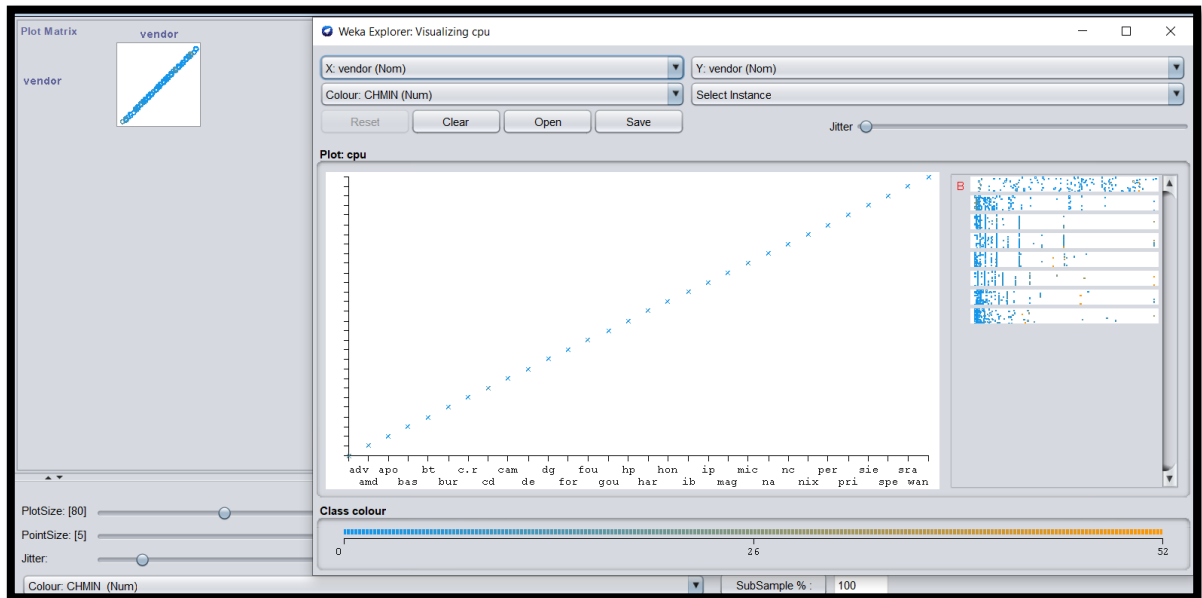
Attribute Subset Evaluator (supervised, Class (numeric): 6 CHMIN):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 1,5,7,8 : 4
                    vendor

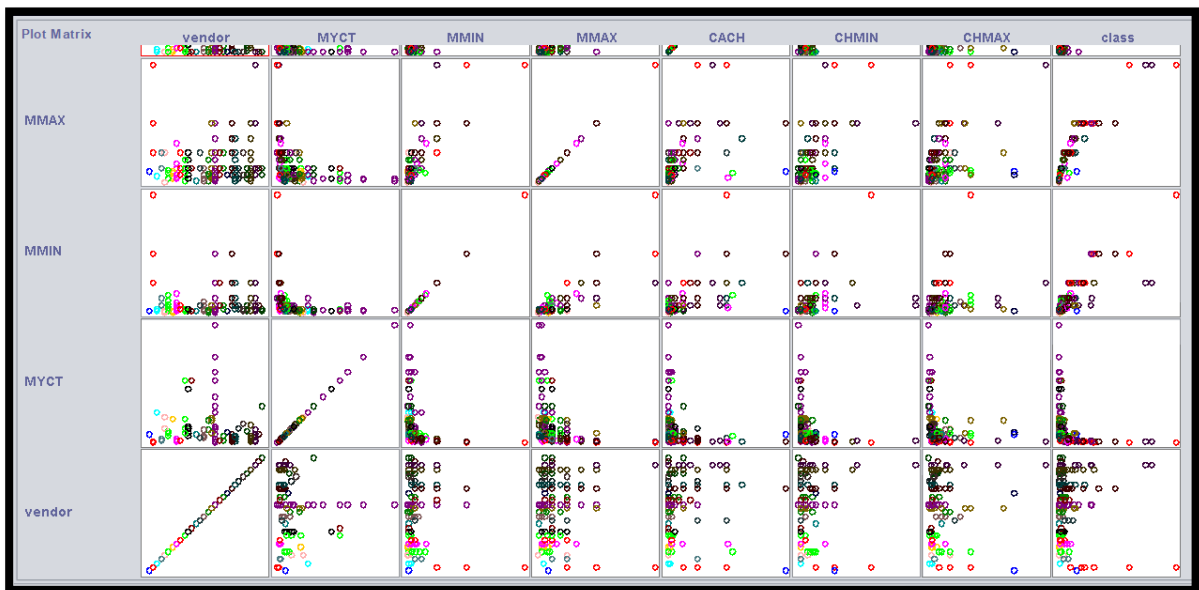
```

The 'Result list (right-click for options)' section shows two entries: '00:37:04 - BestFirst + CfsSubsetEval' and '00:40:42 - BestFirst + CfsSubsetEval'.

- Se puede ver en la grafica, las datos del vendedor con el porcentaje de la respectiva marca.



- Grafica con los datos del vendedor y también de sus propiedades. Gráficas 2D que relacionan pares de atributos.



- **Conclusiones:**
 - Weka nos ayuda a realizar minería de datos, y aprendizaje automático, presenta varios algoritmos aprobados para poder realizar entrenamiento.