

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

BUSINESS INTELLIGENCE

Nombre: Cunduri Oscar

EXÁMEN 2

**1) A continuación se presenta una lista de transacciones con productos A; B;C;D;E;F;G para un almacén. Encontrar el conjunto de ítems frecuentes con 75% de soporte mínimo. Utilizar el algoritmo A priori para responder esta pregunta e indicar los pasos realizados manualmente. Indicar dos reglas de asociación con su soporte y confianza, no utilizar librerías de software para resolver el problema .(5 puntos)**

1 A B C E F G

2 A B E F G

3 A B C E F

4 B C F G

5 A B C E F G

6 A C F G

7 A B C D F

8 D E F G

9 A B C D E

10 A B C F G

a) Soporte individual

Soporte de A:  $8/10 = 0.80$

Soporte de B:  $8/10 = 0.80$

Soporte de C:  $8/10 = 0.80$

Soporte de D:  $3/10 = 0.30$

Soporte de E:  $6/10 = 0.60$

Soporte de F:  $9/10 = 0.90$

Soporte de G:  $7/10 = 0.70$

b) Conjunto de ítems frecuentes

Conjunto	Soporte
A, B	0.7
A, C	0.7
A, F	0.7
B, C	0.7
B, F	0.7

C, F	0.7
A, B, C	0.6
A, B, F	0.6
A, C, F	0.6
B, C, F	0.6

c) Reglas de asociación

- Regla:  $\{A, B\} \Rightarrow \{C, F\}$

Soporte: 0.6 (60%)

Confianza: 0.857 (85.7%)

- Regla:  $\{A, C, F\} \Rightarrow \{B\}$

Soporte: 0.6 (60%)

Confianza: 0.857 (85.7%)

**2) Se requiere predecir el precio de un equipo electrónico en función de características tales su potencia, el número de parlantes, el volumen del equipo, su peso . Indicar cómo resolvería el problema, qué datos necesita y simular la solución de un problema con 10 tuplas usando weka. ( 5 puntos)**

Primero definimos el modelo de entrenamiento y test:

@relation PrecioEquipoElectronico

@attribute Potencia numeric

@attribute NumeroParlantes numeric

@attribute Volumen numeric

@attribute Peso numeric

@attribute Precio numeric

@data

100, 2, 20, 5, 350

120, 4, 25, 6, 450

90, 1, 15, 4, 300

110, 3, 22, 5, 400

105, 2, 18, 4.5, 370

95, 2, 21, 5.2, 320

115, 3, 19, 4.7, 380

125, 4, 24, 6.2, 420

80, 1, 14, 3.5, 270

130, 4, 26, 6.5, 460

%

%

%

@relation PrecioEquipoElectronico

@attribute Potencia numeric

@attribute NumeroParlantes numeric

@attribute Volumen numeric

@attribute Peso numeric

@attribute Precio numeric

@data

95, 2, 18, 4.5, ?

%

%

%

Ahora se entrenaría al modelo en weka con datos de entrenamiento, con el fin de encontrar la fórmula o modelo que predice el precio.

Start

Stop

Result list (right-click for options)

08:21:04 - functions.LinearRegression

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

Linear Regression Model

Precio =

3.2495 \* Potencia +

12.9055 \* Volumen +

-43.191 \* Peso +

-20.4173

Time taken to build model: 0.1 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9405
Mean absolute error	18.4522
Root mean squared error	21.5977
Relative absolute error	32.804 %
Root relative squared error	32.909 %
Total Number of Instances	10

Status

Finalmente, usamos este modelo para predecir el precio para la tupla que se usará para test.

Start

Stop

Result list (right-click for options)

08:21:04 - functions.LinearRegression

08:23:15 - functions.LinearRegression

Re-evaluation on test set

User supplied test set

Relation: PrecioEquipoElectronico

Instances: unknown (yet). Reading incrementally

Attributes: 5

=== Summary ===

Total Number of Instances	0
Ignored Class Unknown Instances	1

=== Re-evaluation on test set ===

User supplied test set

Relation: PrecioEquipoElectronico

Instances: unknown (yet). Reading incrementally

Attributes: 5

=== Predictions on user test set ===

inst#	actual	predicted	error
1	?	326.22	?

=== Summary ===