

# INSTITUTO POLITÉCNICO NACIONAL ESCUELA SUPERIOR DE CÓMPUTO

Unidad de Aprendizaje: Minería de datos

Ciclo escolar: 2023–2

## **Proyecto No. 3. Reglas de Asociación**

Grupo: 3CV15

Equipo: 4

Nombre de los integrantes del equipo:

- 1) Flores Ponce Alan Marcelo  
García Cruz Octavio Arturo
- 2) Sampayo Hernández Mauro
- 3)



Instituto Politécnico Nacional



Escuela Superior de Cómputo

Data Mining

Fabiola Ocampo Botello

Generación de Reglas Asociación

Equipo #4

- Sampayo Hernández Mauro
- García Cruz Octavio Arturo
- Flores Ponce Alan Marcelo

3CM5

## PARTE 1. Ejercicio descriptivo de la generación de reglas de asociación

Realice el proceso (paso a paso) para generar las reglas de asociación del ejercicio descrito y los parámetros:  $F_i = 2$ .  $\text{min.sup} = 2$ .  $\text{min.conf} = 75\%$  (modificable)  
 Genere un conjunto de datos con 7 productos y 10 transacciones (2 con dos productos, 1 de tres productos, 4 de cuatro productos, 2 de cinco productos y 1 de seis productos).

Agregue los siguientes apartados:

### 1) Introducción

El conjunto de datos que será utilizado será el que se muestra a continuación, el cual representa la venta de electrodomésticos:

$I = \{\text{Televisor, Equipo de Sonido, Alexa, Consola de Videojuegos, Impresora, Laptop, Reproductor de DVD}\}$

### 2) Desarrollo

#### a. Listado de transacciones

No. Transacción	Ítems comprados
1	Televisor, Reproductor de DVD
2	Laptop, Impresora
3	Televisor, Equipo de sonidos, Consola de Videojuegos
4	Televisor, Equipo de sonidos, Consola de Videojuegos, Alexa
5	Reproductor de DVD, Impresora, Laptop, Equipo de Sonido
6	Impresora, Equipo de Sonido, Alexa, Consola de Videojuegos
7	Televisor, Laptop, Reproductor de DVD, Equipo de Sonido
8	Televisor, Equipo de Sonido, Alexa, Consola de Videojuegos, Impresora
9	Alexa, Consola de Videojuegos, Impresora, Laptop, Reproductor de DVD
10	Televisor, Equipo de Sonido, Alexa, Consola de Videojuegos, Impresora, Laptop

Transacción	Televisor	Equipo de Sonido	Alexa	Consola de Videojuegos	Impresora	Laptop	Reproductor de DVD
1	1	0	0	0	0	0	1
2	0	0	0	0	1	1	0
3	1	1	0	1	0	0	0
4	1	1	1	1	0	0	0
5	0	1	0	0	1	1	1
6	0	1	1	1	1	0	0
7	1	1	0	0	0	1	1

<b>8</b>	1	1	1	1	1	0	0
<b>9</b>	0	0	1	1	1	1	1
<b>10</b>	1	1	1	1	1	1	0
<b>Suma</b>	<b>6</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>6</b>	<b>5</b>	<b>4</b>

b. Representación tabular

3) Primera Etapa

C1:

<b>Itemset</b>	<b>Soporte</b>
Televisor	6
Equipo de Sonido	7
Alexa	5
Consola de Videojuegos	6
Impresora	6
Laptop	5
Reproductor de DVD	4

Se analiza si las ocurrencias cumplen con  $\text{min\_sup} = 2$ .  $\text{min\_sup} \geq \text{fi}() = 2$ .  
Todos cumplen. Por lo tanto, el resultado es:

L1:

<b>Itemset</b>	<b>Soporte</b>
Televisor	6
Equipo de Sonido	7
Alexa	5
Consola de Videojuegos	6
Impresora	6
Laptop	5
Reproductor de DVD	4

4) Segunda Etapa

Se generan los 2-itemset candidatos para el conjunto C2, lo cual se realiza a partir de L1.

<b>Itemset</b>	<b>Soporte</b>
Televisor	6
Equipo de Sonido	7
Alexa	5
Consola de Videojuegos	6
Impresora	6
Laptop	5

Reproductor de DVD	4
--------------------	---

Itemset	Soporte
Televisor	6
Equipo de Sonido	7
Alexa	5
Consola de Videojuegos	6
Impresora	6
Laptop	5
Reproductor de DVD	4

Itemset	Soporte
{Televisor, Equipo de Sonido}	5
{Televisor, Alexa}	3
{Televisor, Consola de Videojuegos }	4
{Televisor, Impresora }	2
{Televisor, Laptop }	2
{Televisor, Reproductor de DVD }	2
{Equipo de Sonido, Alexa}	4
{Equipo de Sonido, Consola de Videojuegos}	5
{Equipo de Sonido, Impresora }	4
{Equipo de Sonido, Laptop }	3
{Equipo de Sonido, Reproductor de DVD }	2
{Alexa, Consola de videojuegos}	5
{Alexa, Impresora}	4
{Alexa, Laptop}	2
<b>{Alexa, Reproductor de DVD}</b>	<b>1</b>
{Consola de Videojuegos, Impresora}	4
{Consola de Videojuegos, Laptop }	2
<b>{Consola de Videojuegos, Reproductor de DVD }</b>	<b>1</b>
{Impresora, Laptop }	4
{Impresora, Reproductor de DVD }	2
{ Laptop , Reproductor de DVD }	3

Se analiza si las ocurrencias cumplen con  $\text{min\_sup} = 2$ .  $\text{min\_sup} \geq f_i()$  = 2.

L2:

<b>Itemset</b>	<b>Soporte</b>
{Televisor, Equipo de Sonido}	5
{Televisor, Alexa}	3
{Televisor, Consola de Videojuegos }	4
{Televisor, Impresora }	2
{Televisor, Laptop }	2
{Televisor, Reproductor de DVD }	2
{Equipo de Sonido, Alexa}	4
{Equipo de Sonido, Consola de Videojuegos}	5
{Equipo de Sonido, Impresora }	4
{Equipo de Sonido, Laptop }	3
{Equipo de Sonido, Reproductor de DVD }	2
{Alexa, Consola de videojuegos}	5
{Alexa, Impresora}	4
{Alexa, Laptop}	2
{Consola de Videojuegos, Impresora}	4
{Consola de Videojuegos, Laptop }	2
{Impresora, Laptop }	4
{Impresora, Reproductor de DVD }	2
{ Laptop , Reproductor de DVD }	3

### **Tercera Iteración**

<b>Itemset</b>	<b>Soporte</b>
{Televisor, Equipo de Sonido}	5
{Televisor, Alexa}	3
{Televisor, Consola de Videojuegos }	4
{Televisor, Impresora }	2
{Televisor, Laptop }	2
{Televisor, Reproductor de DVD }	2
{Equipo de Sonido, Alexa}	4
{Equipo de Sonido, Consola de Videojuegos}	5
{Equipo de Sonido, Impresora }	4
{Equipo de Sonido, Laptop }	3
{Equipo de Sonido, Reproductor de DVD }	2
{Alexa, Consola de videojuegos}	5
{Alexa, Impresora}	4

{Alexa, Laptop}	2
{Consola de Videojuegos, Impresora}	4
{Consola de Videojuegos, Laptop }	2
{Impresora, Laptop }	4
{Impresora, Reproductor de DVD }	2
{ Laptop , Reproductor de DVD }	3

<b>Itemset</b>	<b>Soporte</b>
{Televisor, Equipo de Sonido, Alexa}	3
{Televisor, Equipo de Sonido, Consola de Videojuegos }	4
{Televisor, Equipo de Sonido, Impresora}	2
{Televisor, Laptop, Reproductor de DVD}	1
{Televisor, Equipo de Sonido, Laptop}	2
{Televisor, Equipo de Sonido, Reproductor de DVD}	1
{Televisor, Alexa, Consola de videojuegos}	4
{Televisor, Alexa, Impresora}	2
{Televisor, Alexa, Laptop}	1
{Televisor, Alexa, Reproductor de DVD}	0
{Televisor, consola de videojuegos, impresora}	2
{Televisor, Consola de videojuegos, Laptop}	1
{Televisor, Consola de videojuegos, Reproductor de DVD}	0
{Televisor, Impresora, Laptop}	1
{Televisor, Impresora, Reproductor de DVD}	0
{Equipo de Sonido, Alexa, Consola de Videojuegos}	4
{Equipo de Sonido, Alexa, Impresora}	3
{Equipo de Sonido, Alexa, Laptop}	1
{Equipo de Sonido, Alexa, Reproductor de DVD}	0
{Equipo de Sonido, Consola de Videojuegos, Impresora}	3
{Equipo de Sonido, Consola de Videojuegos, Laptop}	1
{Equipo de Sonido, Consola de Videojuegos, Reproductor de DVD}	0

{Equipo de Sonido, Impresora, Laptop}	2
{Equipo de Sonido, Impresora, Reproductor de DVD}	1
{Equipo de Sonido, Laptop, Reproductor de DVD}	2
{Alexa, Consola de videojuegos, Impresora}	4
{Alexa, Consola de videojuegos, Laptop}	2
{Alexa, Consola de videojuegos, Reproductor de DVD}	1
{Alexa, Impresora, Laptop}	2
{Alexa, Impresora, Reproductor de DVD}	1
{Alexa, Laptop, Reproductor de DVD}	1
{Consola de videojuegos, Impresora, Laptop}	2
{Consola de videojuegos, Impresora, Reproductor de DVD}	1
{Impresora, Laptop, Reproductor de DVD}	1

Se analiza si las ocurrencias cumplen con  $\min\_sup = 2$ .  $\min\_sup \geq fi() = 2$ .

### L3:

Itemset	Soporte
{Televisor, Equipo de Sonido, Alexa}	3
{Televisor, Equipo de Sonido, Consola de Videojuegos}	4
{Televisor, Equipo de Sonido, Impresora}	2
{Televisor, Equipo de Sonido, Laptop}	2
{Televisor, Alexa, Consola de videojuegos}	4
{Televisor, Alexa, Impresora}	2
{Televisor, consola de videojuegos, impresora}	2
{Equipo de Sonido, Alexa, Consola de Videojuegos}	4
{Equipo de Sonido, Alexa, Impresora}	3
{Equipo de Sonido, Consola de Videojuegos, Impresora}	3
{Equipo de Sonido, Impresora, Laptop}	2



{Equipo de Sonido, Impresora, Reproductor de DVD}	1
{Equipo de Sonido, Laptop, Reproductor de DVD}	2
{Alexa, Consola de videojuegos, Impresora}	4
{Alexa, Consola de videojuegos, Laptop}	2
{Alexa, Impresora, Laptop}	2
{Consola de videojuegos, Impresora, Laptop}	2

#### Cuarta Iteración

Itemset	Soporte
{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos}	3
{Televisor, Equipo de Sonido, Alexa, Impresora}	2
{Televisor, Equipo de Sonido, Alexa, Laptop}	1
{Televisor, Equipo de Sonido, Alexa, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Consola de Videojuegos, Impresora}	2
{Televisor, Equipo de Sonido, Consola de Videojuegos, Laptop}	1
{Televisor, Equipo de Sonido, Consola de Videojuegos, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Impresora, Laptop}	1
{Televisor, Equipo de Sonido, Impresora, Reproductor de DVD}	0
{Televisor, Alexa, Consola de videojuegos, Impresora}	2
{Televisor, Alexa, Consola de videojuegos, Laptop}	1
{Televisor, Alexa, Consola de videojuegos, Reproductor de DVD}	0
{Televisor, consola de videojuegos, impresora, Laptop}	1
{Televisor, consola de videojuegos, impresora, Reproductor de DVD}	0

{Equipo de Sonido, Alexa, Consola de Videojuegos, Impresora}	3
{Equipo de Sonido, Alexa, Consola de Videojuegos, Laptop}	1
{Equipo de Sonido, Alexa, Consola de Videojuegos, Reproductor de DVD}	0
{Equipo de Sonido, Alexa, Impresora, Laptop}	1
{Equipo de Sonido, Alexa, Impresora, Reproductor de DVD}	0
{Equipo de Sonido, Consola de Videojuegos, Impresora, Laptop}	1
{Equipo de Sonido, Consola de Videojuegos, Impresora, Reproductor de DVD}	0
{Equipo de Sonido, Impresora, Laptop, Reproductor de DVD}	1
{Alexa, Consola de videojuegos, Impresora, Laptop}	2
{Alexa, Consola de videojuegos, Impresora, Reproductor de DVD}	1
{Alexa, Consola de videojuegos, Laptop, Reproductor de DVD}	1
{Alexa, Impresora, Laptop, Reproductor de DVD}	1
{Consola de videojuegos, Impresora, Laptop, Reproductor de DVD}	1

Se analiza si las ocurrencias cumplen con  $\min\_sup = 2$ .  $\min\_sup \geq fi() = 2$ .

#### L4:

Itemset	Soporte
{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos}	3
{Televisor, Equipo de Sonido, Alexa, Impresora}	2
{Televisor, Equipo de Sonido, Consola de Videojuegos, Impresora}	2
{Televisor, Alexa, Consola de videojuegos, Impresora}	2
{Equipo de Sonido, Alexa, Consola de Videojuegos, Impresora}	3
{Alexa, Consola de videojuegos, Impresora, Laptop}	2

#### Quinta Iteración

Itemset	Soporte
{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos, Impresora}	2

{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos, Laptop}	1
{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Alexa, Impresora, Laptop}	1
{Televisor, Equipo de Sonido, Alexa, Impresora, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Alexa, Laptop, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Consola de videojuegos, Impresora, Laptop}	1
{Televisor, Equipo de Sonido, Consola de videojuegos, Impresora, Reproductor de DVD}	0
{Televisor, Equipo de Sonido, Impresora, Laptop, Reproductor de DVD}	0
{Televisor, Alexa, Consola de videojuegos, Impresora, Laptop}	1
{Televisor, Alexa, Consola de videojuegos, Impresora, Reproductor de DVD}	0
{Televisor, Alexa, Impresora, Laptop Reproductor de DVD}	0
{Televisor, Consola de videojuegos, Impresora, Laptop, Reproductor de DVD}	0
{Equipo de sonido, Alexa, Consola de videojuegos, Impresora, Laptop}	1
{Equipo de sonido, Alexa, Consola de videojuegos, Laptop, Reproductor de DVD}	0
{Equipo de sonido, Alexa, Impresora, Laptop, Reproductor de DVD}	0
{Equipo de sonido, Consola de videojuegos, Impresora, Laptop, Reproductor de DVD}	0
{Alexa, Consola de videojuegos, Impresora, Laptop, Reproductor de DVD}	1

**L5:**

<b>Itemset</b>	<b>Soporte</b>
{Televisor, Equipo de Sonido, Alexa, Consola de videojuegos, Impresora}	2

5) Evaluación de las reglas

6) Conclusiones



Instituto Politécnico Nacional



Escuela Superior de Cómputo

Data Mining

Fabiola Ocampo Botello

Aprendizaje Supervisado

Equipo #4

- Sampayo Hernández Mauro
- García Cruz Octavio Arturo
- Flores Ponce Alan Marcelo

3CM5

## PARTE 2. Aprendizaje supervisado

Considere el siguiente conjunto de datos:

Nombre:	Conjunto de datos de registros de votación del Congreso ( <i>Congressional Voting Records Data Set</i> )
Objetivo:	Este conjunto de datos incluye los votos de cada uno de los congresistas de la Cámara de Representantes de los Estados Unidos sobre los 16 votos clave identificados por la CQA. La CQA enumera nueve tipos diferentes de votos: votado a favor, emparejado a favor y anunciado a favor (estos tres simplificados a sí), votado en contra, emparejado en contra y anunciado en contra (estos tres simplificados a no), votado presente, votado presente para evitar conflicto de intereses, y no votó ni dio a conocer una posición (estos tres simplificados a una disposición desconocida).
Créditos:	Congressional Quarterly Almanac, 98th Congress, 2nd session 1984, Volume XL: Congressional Quarterly Inc. Washington, D.C., 1985. Donante: Jeff Schlimmer (Jeffrey.Schlimmer '@' a.gp.cs.cmu.edu)
Enlace de acceso:	<a href="https://archive.ics.uci.edu/ml/datasets/congressional+voting+records">https://archive.ics.uci.edu/ml/datasets/congressional+voting+records</a>
Clases:	Demócrata, Republicano

1) Crear el diccionario de datos con los siguientes campos:

Nombre	Significado	Tipo	Dominio
Nombre de la clase	Describe a que partido político pertenece	Categórico	Demócrata, Republicano
Niños discapacitados	Apoyo a Niños Discapacitados	Categórico	Yes, No
Costos compartidos del proyecto de agua	Se refiere a los gastos compartidos entre diferentes entidades	Categórico	Yes, No
Adopción de la resolución presupuestaria	A favor o en contra de la resolución presupuestaria	Categórico	Yes, No
Congelación de honorarios médicos	Apoya la Congelación de honorarios médicos	Categórico	Yes, No
Ayuda-el-salvador	Propone ayuda a la gente Salvadoreña	Categórico	Yes, No
Grupos-religiosos-en-escuelas	Propone grupos Religiosos en las escuelas	Categórico	Yes, No

<b>Prohibición de pruebas anti-satélite</b>	Esta a favor de la Prohibición de pruebas anti-satélite	Categorico	Yes, No
<b>Ayuda-a-los-contras-nicaraguenses</b>	Propone ayuda a las gentes Nicaraguenses	Categorico	Yes, No
<b>mx-misil</b>	A favor o en contra de mx-misil	Categorico	Yes, No
<b>inmigración</b>	Apoya o esta en contra de la inmigración hacia Estados Unidos	Categorico	Yes, No
<b>synfuels-corporation-cutback</b>	Apoya las synfuels-corporation-cutback	Categorico	Yes, No
<b>gasto en educación</b>	Prone reducir el gasto en la educacion	Categorico	Yes, No
<b>superfund-right-to-sue</b>	Apoya las superfund-right-to-sue	Categorico	Yes, No
<b>crimen</b>	Propuestas para la reducción del crimen	Categorico	Yes, No
<b>exportaciones libres de impuestos</b>	Apoya las exportaciones libres de impuestos	Categorico	Yes, No
<b>Ley de administración de exportaciones-Sudáfrica</b>	A favor o en contra de la Ley de administración de exportaciones-Sudáfrica	Categorico	Yes, No

- 2) Realice el tratamiento de datos necesario para aplicar la técnica de minería de datos de reglas de asociación. (Ustedes establezcan la configuración de los parámetros para la generación de las reglas).

Primeramente, se carga el archivo con extensión .data

Dialog - 3:1 - File Reader (cargaDatos)

File

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables Job Manager Selection Memory Policy

Input location

Read from Local File System

Mode ☒ File ☐ Files in folder

File C:\Users\mauro\Documents\Pasar a IPN\_ESCOM\Data Mining\Practicas y Tareas\Tarea\_reglasAsociacion\archivos\house-votes-84.data Browse...

Reader options

Format

Autodetect format

Column delimiter , Row delimiter ☒ Line break ☐ Custom \r\n

Quote char " Quote escape char \

Comment char #

Has column header ☐ Has row ID ☐

Support short data rows ☐ Prepend file index to row ID ☐

Preview

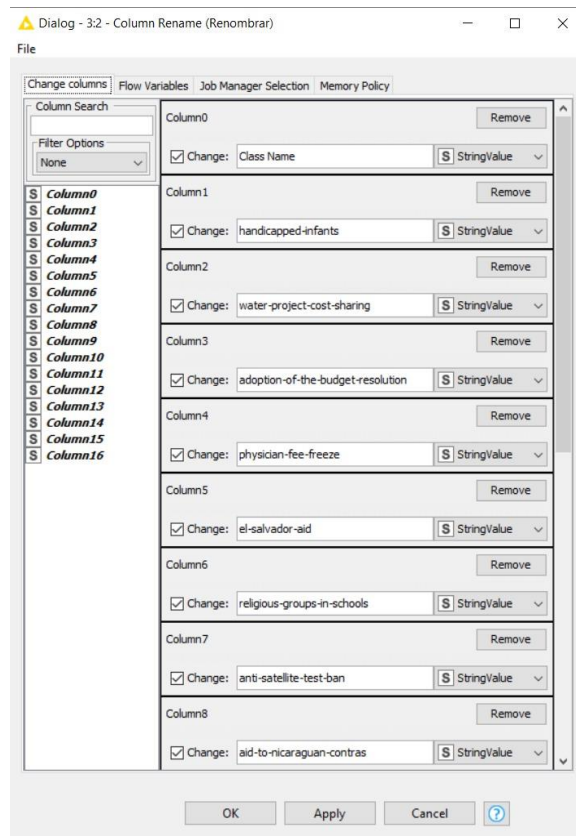
The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	Column0	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11
Row0	republican	n	y	n	y	y	y	n	n	n	y	?
Row1	republican	n	y	n	y	y	y	n	n	n	n	n
Row2	democrat	?	y	y	?	y	y	n	n	n	n	y
Row3	democrat	n	y	y	n	?	y	n	n	n	n	y
Row4	democrat	y	y	y	n	y	y	n	n	n	n	y
Row5	democrat	n	y	y	n	y	y	n	n	n	n	n
Row6	democrat	n	y	n	y	y	y	n	n	n	n	n
Row7	republican	n	y	n	y	y	y	n	n	n	n	n
Row8	republican	n	y	n	y	y	y	n	n	n	n	n
Row9	democrat	y	y	y	n	n	n	y	y	y	n	n

OK Apply Cancel ?



Al tener nuestros datos cargados procederemos a cambiar el nombre las columnas con un Column Rename.



Una vez completado el tratamiento anterior procedemos a conectar nuestro column rename a un nodo de python el cual tendra como objetivo agregar a cada dato la palabra 'yes to' o 'no to' dependiendo si tiene establecido 'y' o 'n' por default, y a su vez agregara el nombre de la columna al que pertenece.

```
import knime.scripting.io as knio

# Conversion de KNIME a DataFrame
votos = knio.input_tables[0].to_pandas()

# Sustitucion de los valores de todas las columnas, a valores mas comprensibles
for col in votos.columns:
    votos[col]=votos[col].replace('y', 'yes to ' + col)
    votos[col]=votos[col].replace('n', 'no to ' + col)
    votos[col]=votos[col].replace('?', 'abstention to ' + col)
```

```
# Tabla de Salida
```

```
knio.output_tables[0] = knio.Table.from_pandas(votos)
```

- 3) Elija las reglas de asociación que tengan como consecuente las clases indicadas (Demócrata, Republicano) para identificar las características de cada partido político.

Se utilizo un **Soporte** del 15% y una **Confianza** del 60%, para generar las reglas de asociación.

### Demócratas

Row ID	S Conseq...	S Antecedent
Row290_1	democrat	yes to synfuels-corporation-cutback
Row290_2	democrat	no to superfund-right-to-sue
Row290_3	democrat	yes to aid-to-nicaraguan-contras
Row290_4	democrat	no to physician-fee-freeze
Row414_2	democrat	no to el-salvador-aid
Row414_3	democrat	no to education-spending
Row454_3	democrat	yes to anti-satellite-test-ban
Row472_4	democrat	yes to adoption-of-the-budget-resolution
Row631_4	democrat	yes to export-administration-act-south-africa
Row1826_1	democrat	no to religious-groups-in-schools
Row1826_2	democrat	no to crime
Row1826_3	democrat	yes to duty-free-exports
Row1826_5	democrat	yes to mx-missile
Row6635_3	democrat	yes to handicapped-infants
Row13606_5	democrat	no to immigration
Row16494_7	democrat	no to synfuels-corporation-cutback
Row34916_2	democrat	no to water-project-cost-sharing
Row133606_2	democrat	yes to immigration
Row214969_1	democrat	yes to water-project-cost-sharing
Row226766_2	democrat	no to handicapped-infants
Row228537_4	democrat	yes to religious-groups-in-schools

## Republicanos

Row ID	S Conseq...	S Antecedent
Row46217_1	republican	no to adoption-of-the-budget-resolution
Row46217_2	republican	yes to education-spending
Row46217_3	republican	yes to physician-fee-freeze
Row46217_4	republican	no to aid-to-nicaraguan-contras
Row46217_5	republican	no to anti-satellite-test-ban
Row46217_6	republican	no to mx-missile
Row46217_7	republican	yes to superfund-right-to-sue
Row46217_8	republican	yes to el-salvador-aid
Row46217_9	republican	no to duty-free-exports
Row46227_9	republican	no to handicapped-infants
Row46237_9	republican	yes to crime
Row46247_9	republican	no to synfuels-corporation-cutback
Row46257_9	republican	yes to religious-groups-in-schools
Row60811_5	republican	yes to export-administration-act-south-africa
Row66842_3	republican	yes to water-project-cost-sharing
Row67396_4	republican	no to immigration
Row67414_5	republican	yes to immigration
Row96857_2	republican	no to water-project-cost-sharing

- 4) Describa las características generadas por las reglas de asociación que distinguen cada partido.

Partido	Características (listado)
Demócrata	<ul style="list-style-type: none"> <li>– <b>EN CONTRA</b> de la ley <i>superfund right to sue</i>.</li> <li>– <b>EN CONTRA</b> de la <i>congelación de honorarios médicos</i>.</li> <li>– <b>EN CONTRA</b> de la <i>ayuda a El Salvador</i>.</li> <li>– <b>EN CONTRA</b> del <i>gasto en la educación</i>.</li> <li>– <b>EN CONTRA</b> de <i>tomar medidas para combatir el crimen</i>.</li> <li>– <b>A FAVOR</b> de la <i>ayuda a los contras nicaragüenses</i>.</li> <li>– <b>A FAVOR</b> de la <i>prohibición de pruebas antisatélite</i>.</li> <li>– <b>A FAVOR</b> de la <i>adopción de la resolución presupuestaria</i>.</li> <li>– <b>A FAVOR</b> de la <i>ley de administración de exportación a Sudáfrica</i>. *</li> <li>– <b>A FAVOR</b> de las <i>exportaciones libres de impuestos</i>.</li> <li>– <b>A FAVOR</b> de los <i>misiles-mx</i>.</li> <li>– Existe una <b>DIVISIÓN MARCADA</b> aparente, entre los simpatizantes de este partido, que están <b>a favor</b> y <b>en contra</b> de la <i>inmigración</i>, la <i>reducción de ingresos destinados a la Synthetic Fuels Corporation</i>, el <i>proyecto de costos compartidos de agua</i>, el <i>apoyo a los niños discapacitados</i> y la <i>presencia de grupos religiosos en escuelas</i>.</li> </ul>
Republicano	<ul style="list-style-type: none"> <li>– <b>EN CONTRA</b> de la <i>adopción de la resolución presupuestaria</i>.</li> </ul>

	<ul style="list-style-type: none"> <li>– <b>EN CONTRA</b> de la <i>ayuda a los contras nicaragüenses.</i></li> <li>– <b>EN CONTRA</b> de la <i>prohibición de pruebas antisatélite.</i></li> <li>– <b>EN CONTRA</b> de los <i>misiles-mx.</i></li> <li>– <b>EN CONTRA</b> de las <i>exportaciones libres de impuestos.</i></li> <li>– <b>EN CONTRA</b> del <i>apoyo a los niños discapacitados.</i></li> <li>– <b>EN CONTRA</b> de la <i>reducción de ingresos destinados a la Synthetic Fuels Corporation.</i></li> <li>– <b>A FAVOR</b> del <i>gasto en la educación.</i></li> <li>– <b>A FAVOR</b> de la <i>congelación de honorarios médicos.</i></li> <li>– <b>A FAVOR</b> de la <i>ley superfund right to sue.</i></li> <li>– <b>A FAVOR</b> de la <i>ayuda a El Salvador.</i></li> <li>– <b>A FAVOR</b> de <i>tomar medidas para combatir el crimen.</i></li> <li>– <b>A FAVOR</b> de la <i>presencia de grupos religiosos en escuelas.</i></li> <li>– <b>A FAVOR</b> de la <i>ley de administración de exportación a Sudáfrica. *</i></li> <li>– Existe una <b>DIVISIÓN MARCADA</b> aparente, entre los simpatizantes de este partido, que están <b>a favor</b> y <b>en contra</b> de la <i>inmigración y el proyecto de costos compartidos de agua.</i></li> </ul>
--	---

Agregue las reglas de asociación que generó tales características



Instituto Politécnico Nacional



Escuela Superior de Cómputo

Data Mining

Fabiola Ocampo Botello

Aprendizaje No Supervisado

Equipo #4

- Sampayo Hernández Mauro
- García Cruz Octavio Arturo
- Flores Ponce Alan Marcelo

3CM5

### PARTE 3. Aprendizaje No supervisado

Crear un reporte (tipo catálogo) de ofertas mediante la herramienta BIRT de KNIME que sugiera productos de calzado. Debe generar reglas de asociación.

Diseñe un catálogo recomendando los productos de calzado más vendidos o menos vendidos (ustedes determinarán dicho criterio). Los datos mostrados en el catálogo los generará considerando las reglas generadas:

Antecedente → Consecuente

En el reporte mostrará los nombres y los precios, marca (puede incluir algún otro dato) de los productos tanto de antecedente como de consecuente y con base en el significado de soporte, confianza, lift o algún otro parámetro que calcule, realice una oferta de esa asociación.

Elija un conjunto de datos y presente los siguientes datos:

Nombre:	Precios de calzado para mujer. Una lista de 10,000 zapatos para mujer y los precios a los que se venden.
Objetivo:	Este conjunto de datos es una lista de 10,000 zapatos de mujer y la información de los productos actualizada de la base de datos de productos de Datafiniti entre enero de 2018 y octubre de 2018. El conjunto de datos incluye el nombre del zapato, la marca, el precio y más. Cada zapato tendrá una entrada por cada precio que se encuentre y algunos zapatos pueden tener varias entradas.
Créditos:	Esta es una lista de 10,000 zapatos de mujer y la información de sus productos proporcionada por la base de datos de productos de Datafiniti.
Enlace de acceso:	<a href="https://www.kaggle.com/datasets/datafiniti/womens-shoes-prices?select=Datafiniti_Womens_Shoes.csv">https://www.kaggle.com/datasets/datafiniti/womens-shoes-prices?select=Datafiniti_Womens_Shoes.csv</a> Elegir el conjunto de datos: Datafiniti_Womens_Shoes.csv Este es el archivo que contiene más imágenes. Esquema de datos Un esquema completo de los datos está disponible en nuestra documentación de soporte. <a href="https://developer.datafiniti.co/docs/product-data-schema">https://developer.datafiniti.co/docs/product-data-schema</a>

1) Crear el diccionario de datos con los siguientes campos:

Nombre	Significado	Tipo	Dominio
assins	Identificadores de amazon, palabras clave	Palabra clave	?, B016GSDNVM
brand	Marca del producto como los presenta la	Texto	2 lips too Adidas Adidas outdoor Altra

	fuelle de información		. . . . Vans
categories	Lista de palabras clave categóricas usadas por los productos a través de múltiples fuentes	Texto	Clothing, Shoes, Baby & kids, Girls's boots, women shoes, All womens, etc.
colors	Colores del calzado disponibles para el producto	Texto	Angora, Light porcelain blue, aquifer, black, beige, etc.
dimensión	Longitud, anchura y altura del producto. Unidades incluidas	Palabra clave	4-10in x 3.5-11in x 3.94-12.5in
ean	Códigos de barras para los productos. Puede haber más de un código cuando un producto tiene distintas variaciones	Palabra clave	1 – 8 E 10 – 11
imageURLS	Lista de direcciones URL de imagen para este producto	Palabra clave	"www.NombreDeDominio1.com/..." "www.NombreDeDominio2.com/..." "www.NombreDeDominio3.com/..." "www.NombreDeDominioN.com/..."
keys	Lista de identificadores internos de Datafiniti para cada producto. El campo de claves se	Palabra clave	000000000000 - 999999999999

	utiliza para combinar datos sin procesar de fuentes individuales en el registro maestro de Datafiniti		
manufacturer	Fabricante del producto	Texto	ASICS Brinley Co Dr. Scholl's ELLIE SHOES Eastland Shoe Easy Street Fila Journee Collection Lugz Muk Luks Reliable of Milwaukee Riverberry S-1994-red-6 Silver Lilly Sloggers Soft Ones Spring Footwear
manufacturerNumber	Número de modelo del producto	Palabra clave	0015589605 -S 0016485001 -6 105 - SADDLE_BKPK 11930-BKW 12419bkcc 12615CCCGR 12-129 ... Etc.
name	Nombre del producto	Texto	12831 White Black Skechers Shoe memory Foam 2 Lips Too Too Elle Women's High Heels Sandals 2 Lips Too Too Elle Women's Ankle boots A2 by Aerosoles ChlearSaling Women's Sandals ... Etc.
primaryCategories	Lista de categorías estandarizadas a las que pertenece este producto	Texto	Shoes



sizes	Lista de tallas disponibles del zapato	Numérico	5.5 – 12 5.5 – 12 wide/med 35 – 42 4.5 – 9 mens/women 4.5 – 9 US/UK Small/Med/Large
sourceURLs	Direcciones URL utilizadas para generar datos del producto	Palabra clave	“www.NombreDeDominio1.com/...” “www.NombreDeDominio2.com/...” “www.NombreDeDominio3.com/...” “www.NombreDeDominioN.com/...”
upc	Código UPC del producto. (Puede tener más de uno cuando el producto tiene más de una variación)	Palabra clave	000000000000 - 999999999999
websiteIDs	ID's para el producto dentro del sitio web. Los ID están ligados a un minorista en línea en específico.	Palabra clave	“www.NombreDeDominio1.com/...” “www.NombreDeDominio2.com/...” “www.NombreDeDominio3.com/...” “www.NombreDeDominioN.com/...”
weight	Peso del producto. (Unidades incluidas)	Numérico	7 – 14 oz 1lb 2oz – 1lb 7oz

- 2) Realice el tratamiento de datos necesario para aplicar la técnica de minería de datos de reglas de asociación. (Ustedes establezcan la configuración de los parámetros para la generación de las reglas).  
En el conjunto de datos hay productos que aparecen reportados en varios registros, pero los precios son diferentes.  
Elija una muestra de 500 registros.
- 3) Generar un archivo de transacciones y elegir los criterios para generar las reglas de asociación, por ejemplo: recomendar productos de la misma marca,

### **Tratamiento de datos (500 registros)**

Se van a tomar 500 registros, desde el 120 al 620, estos se agregarán a un archivo manualmente, manteniendo únicamente los siguientes atributos o categorías para una simple generación de las reglas de asociación:

### **Archivo de transacciones y elección de criterios**

El archivo de transacciones se encuentra en el siguiente enlace:

[https://docs.google.com/spreadsheets/d/1RR-fCMxg2ri9qNBJPYCMnWyR\\_OIfKsh/edit?usp=sharing&ouid=103536059312424337576&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1RR-fCMxg2ri9qNBJPYCMnWyR_OIfKsh/edit?usp=sharing&ouid=103536059312424337576&rtpof=true&sd=true)

La elección de criterios fue dada tras la experimentación con los parámetros para asegurar que no surjan tantas reglas de asociación o un conjunto vacío. Para las reglas de asociación de marca los datos son los siguientes:

- Minimum set size: 1
- Minimum rule confidence 30%
- Minimum support 5

Para las reglas de asociación de color los datos son los siguientes:

- Minimum set size: 2
- Minimum rule confidence 30%
- Minimum support 5

### **PARTE 4. Flujo de trabajo en Knime**

Coloque en su canal de trabajo del grupo de MT el archivo generado por el Knime y un integrante por equipo enviará su reporte en formato pdf en la asignación correspondiente.