

### Proyecto No. 3. Reglas de Asociación

Grupo: 3CV9 Equipo: 6

Nombre de los integrantes del equipo:

- 1) Medina Granados Alan Alejandro
- 2) Castro Cruces Jorge Eduardo
- 3) Guzman Gutierrez Manuel

*Inserte una portada indicando el tema que aborda*

#### **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:  $F_i = 2$ . min.sup = 2. 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

Las reglas de asociación se usan usualmente para determinar los hechos que ocurren en común dentro de un determinado conjunto de datos.

Esta información se puede utilizar como base para tomar decisiones sobre marketing como precios promocionales para ciertos productos o dónde ubicar estos dentro del supermercado.

## 2) Desarrollo

### Productos a vender.

$I = \{\text{cereal, cerveza, fruta, leche, pañales, pizza, refresco}\}$

#### a. Listado de transacciones

No.Transacción	Ítems comprados
1	{cerveza, pañales}
2	{pizza, refresco}
3	{cereal, fruta, leche}
4	{cerveza, pañales, pizza, refresco}
5	{cereal, fruta, leche, pañales}
6	{cerveza, cereal, refresco, pizza}
7	{cereal, cerveza, leche, pañales}
8	{cereal, leche, pañales, pizza}
9	{cereal, cerveza, fruta, leche, pizza, refresco}
10	{cereal, fruta, leche, pañales, pizza, refresco}

#### b. Representación tabular

### Tabla de transacciones.

Transacción	Cereal	Cerveza	Fruta	Leche	Pañales	Pizza	Refresco
1	0	1	0	0	1	0	0
2	0	0	0	0	0	1	1
3	1	0	1	1	0	0	0
4	0	1	0	0	1	1	1
5	1	0	1	1	1	0	0

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

### 3) Primera Etapa

1.- Generar los conjuntos de elementos frecuentes.

C1 =

Itemset	Soporte
Cereal	7
Cerveza	5
Fruta	4
Leche	6
Pañales	6
Pizza	6
Refresco	5

Se analiza si las ocurrencias cumplen con  $\text{min\_sup} = 2$ .  $\text{min\_sup} \geq F_i = 2$ .  
 Todos cumplen por lo tanto el resultado es

L1 =

Itemset	Soporte
Cereal	7
Cerveza	5
Fruta	4
Leche	6

Pañales	6
Pizza	6
Refresco	5

Se general los 2-itemset candidatos para el conjunto C2, lo cual se realiza apartir de L1

Itemset	Soporte
Cereal	7
Cerveza	5
Fruta	4
Leche	6
Pañales	6
Pizza	6
Refresco	5



Itemset	Soporte
Cereal	7
Cerveza	5
Fruta	4
Leche	6
Pañales	6
Pizza	6
Refresco	5

Itemset	Soporte
Cerveza, Pañales	3
Cereal, Cerveza	3
Cerveza, Fruta	1
Cerveza, Leche	2
Cerveza, Refresco	3
Cerveza, Pizza	3
Cereal, Pañales	4
Fruta, Pañales	2
Leche, Pañales	4
Pañales, Refresco	2
Pañales, Pizza	3
Cereal, Fruta	4
Cereal, Leche	6
Cereal, Refresco	3
Cereal, Pizza	4
Fruta, Leche	4
Fruta, Refresco	2
Fruta, Pizza	2
Leche, Refresco	2
Leche, Pizza	3
Pizza, Refresco	5

Se generan los 3-ItemSet

Itemset	Soporte
---------	---------

Cereal, Cerveza, Pañales	1
Cerveza, Fruta, Pañales	0
Cerveza, leche, Pañales	1
Cerveza, Pañales, Refresco	1
Cerveza, Pañales, Pizza	1
Cereal, Cerveza, Fruta	1
Cereal, Cerveza, Leche	2
Cereal, Cerveza, Refresco	2
Cereal, Cerveza, Pizza	2
Cereal, Fruta, Pañales	2
Cereal, Leche, Pañales	4
Cereal, Pañales, Refresco	1
Cereal, Pañales, Pizza	1
Fruta, Leche, Pañales	2
Fruta, Pañales, Refresco	1
Fruta, Pañales, Pizza	1
Leche, Pañales, Refresco	1
Fruta, Pañales, Pizza	1
Cereal, Fruta, Leche	4
Cereal, Fruta, Refresco	2
Cereal, Fruta, Pizza	2
Cereal, Leche, Refresco	2
Cereal, Fruta, Pizza	2
Fruta, Leche, Refresco	2
Fruta, Leche, Pizza	2
Fruta, Pizza, Refresco	2

Leche, Pizza, Refresco	2
------------------------	---

L3 =

Itemset	Soporte
Cereal, Cerveza, Leche	2
Cereal, Cerveza, Refresco	2
Cereal, Cerveza, Pizza	2
Cereal, Fruta, Pañales	2
Cereal, Leche, Pañales	4
Fruta, Leche, Pañales	2
Cereal, Fruta, Leche	4
Cereal, Fruta, Refresco	2
Cereal, Fruta, Pizza	2
Cereal, Leche, Refresco	2
Cereal, Fruta, Pizza	2
Fruta, Leche, Refresco	2
Fruta, Leche, Pizza	2
Fruta, Pizza, Refresco	2
Leche, Pizza, Refresco	2

Subconjunto	Frecuencia	Decisión
Cereal, Cerveza, Leche		No se poda
Cereal, Cerveza	3	
Cerveza, Leche	2	
Cereal, Leche	6	

Cereal, Cerveza, Refresco		No se poda
Cereal, Cerveza	3	
Cerveza, Refresco	3	
Cereal, Refresco	3	
Cereal, Cerveza, Pizza		No se poda
Cereal, Cerveza	3	
Cerveza, Pizza	3	
Cereal, Pizza	4	
Cereal, Fruta, Pañales		No se poda
Cereal, Pañales	4	
Fruta, Pañales	2	
Cereal, Fruta	4	
Cereal, Leche, Pañales		No se poda
Cereal, Pañales	4	
Leche, Pañales	4	
Cereal, Leche	6	
Fruta, Leche, Pañales		No se poda
Fruta, Pañales	2	
Leche, Pañales	4	
Fruta, Leche	4	
Cereal, Fruta, Leche		No se poda
Cereal, Fruta	4	
Cereal, Leche	6	
Fruta, Leche	4	
Cereal, Fruta, Refresco		No se poda
Cereal, Fruta	4	



Cereal, Refresco	3	
Fruta, Refresco	2	
Cereal, Fruta, Pizza		No se poda
Cereal, Fruta	4	
Cereal, Pizza	4	
Fruta, Pizza	2	
Cereal, Leche, Refresco		No se poda
Cereal, Leche	6	
Cereal, Refresco	3	
Leche, Refresco	2	
Cereal, Fruta, Pizza		No se poda
Cereal, Fruta	4	
Cereal, Pizza	4	
Fruta, Pizza	2	
Fruta, Leche, Refresco		No se poda
Fruta, Leche	4	
Fruta, Refresco	2	
Leche, Refresco	2	
Fruta, Leche, Pizza		No se poda
Fruta, Leche	4	
Fruta, Pizza	2	
Leche, Pizza	3	
Fruta, Pizza, Refresco		No se poda
Fruta, Refresco	2	
Fruta, Pizza	2	
Pizza, Refresco	5	

Leche, Pizza, Refresco		No se poda
Leche, Refresco	2	
Leche, Pizza	3	
Pizza, Refresco	5	

## Segunda Etapa

Confianza del 75%

Regla	Soporte	Confianza
Cereal, Cerveza => Leche	3/10	2/3
Cerveza, Leche => Cereal	2/10	2/2
Cereal, Leche => Cerveza	6/10	2/6
Cereal, Cerveza => Refresco	3/10	2/3
Cerveza, Refresco => Cereal	3/10	2/3
Cereal, Refresco => Cereal	3/10	2/3
Cereal, Cerveza => Refresco	3/10	2/3
Cerveza, Pizza => Cereal	3/10	2/3
Cereal, Pizza => Cerveza	4/10	2/4
Cereal, Pañales => Fruta	4/10	2/4
Fruta, Pañales => Cereal	2/10	2/2
Cereal, Fruta => Pañales	4/10	2/4
Cereal, Pañales => Leche	4/10	4/4
Leche, Pañales => Cereal	4/10	4/4

Cereal, Leche => Pañales	6/10	4/6
Fruta, Pañales => Leche	2/10	2/2
Leche, Pañales => Fruta	4/10	2/4
Fruta, Leche => Pañales	4/10	2/4
Cereal, Fruta => Leche	4/10	2/4
Cereal, Leche => Fruta	6/10	4/6
Fruta, Leche => Cereal	4/10	4/4
Cereal, Fruta => Refresco	4/10	4/4
Cereal, Refresco => Fruta	3/10	2/3
Fruta, Refresco => Cereal	2/10	2/2
Cereal, Fruta => Pizza	4/10	2/4
Cereal, Pizza => Fruta	4/10	2/4
Fruta, Pizza => Cereal	2/10	2/2
Cereal, Leche => Refresco	6/10	2/6
Cereal, Refresco => Leche	3/10	2/3
Leche, Refresco => Cereal	2/10	2/2
Cereal, Fruta => Pizza	4/10	2/4
Cereal, Pizza => Fruta	4/10	2/4
Fruta, Pizza => Cereal	2/10	2/2
Fruta, Leche => Refresco	4/10	2/4
Fruta, Refresco => Leche	2/10	2/2
Leche, Refresco => Fruta	2/10	2/2

Fruta, Leche => Pizza	4/10	2/4
Fruta, Pizza => Leche	2/10	2/2
Leche, Pizza => Fruta	3/10	2/3
Fruta, Refresco => Pizza	2/10	2/2
Fruta, Pizza => Refresco	2/10	2/2
Pizza, Refresco => Fruta	5/10	2/5
Leche, Refresco => Pizza	2/10	2/2
Leche, Pizza => Refresco	3/10	2/3
Pizza, Refresco => Leche	5/10	2/5

#### Reglas finales

Regla	Soporte	Confianza
Cerveza, Leche => Cereal	2/10	2/2
Fruta, Pañales => Cereal	2/10	2/2
Cereal, Pañales => Leche	4/10	4/4
Leche, Pañales => Cereal	4/10	4/4
Fruta, Pañales => Leche	2/10	2/2
Fruta, Leche => Cereal	4/10	4/4
Cereal, Fruta => Refresco	4/10	4/4
Fruta, Refresco => Cereal	2/10	2/2
Leche, Refresco => Cereal	2/10	2/2
Fruta, Pizza => Cereal	2/10	2/2
Fruta, Refresco => Leche	2/10	2/2
Leche, Refresco => Fruta	2/10	2/2

Fruta, Pizza => Leche	2/10	2/2
Fruta, Refresco => Pizza	2/10	2/2
Fruta, Pizza => Refresco	2/10	2/2
Leche, Refresco => Pizza	2/10	2/2

#### 4) Evaluación de las reglas

Lift = confianza / proporción a priori del consecuente (soporte)

1.- Cerveza, Leche => Cereal

Lift =  $(2/2) / (2/10)$

2.- Fruta, Pañales => Cereal

Lift =  $(2/10)/(2/2)$

3.- Cereal, Pañales => Leche

Lift =  $(4/10)/(4/4)$

4.-Leche, Pañales => Cereal

Lift =  $(4/10)/(4/4)$

5.- Fruta, Pañales => Leche

Lift =  $(2/10)/(2/2)$

6.- Fruta, Leche => Cereal

Lift =  $(4/10)/(4/4)$

7.- Cereal, Fruta => Refresco

Lift =  $(4/10)/(4/4)$

8.- Fruta, Refresco => Cereal

Lift =  $(2/10)/(2/2)$

9.- Leche, Refresco => Cereal

Lift =  $(2/10)/(2/2)$

10.- Fruta, Pizza => Cereal

Lift =  $(2/10)/(2/2)$

11.- Fruta, Refresco => Leche

Lift =  $(2/10)/(2/2)$

12.- Leche, Refresco => Fruta

Lift =  $(2/10)/(2/2)$

13.- Fruta, Pizza => Leche

Lift =  $(2/10)/(2/2)$

14.- Fruta, Refresco => Pizza

Lift =  $(2/10)/(2/2)$

15.- Fruta, Pizza => Refresco

Lift =  $(2/10)/(2/2)$

16.- Leche, Refresco => Pizza

Lift =  $(2/10)/(2/2)$

## 5) Conclusiones

Sí bien las reglas de asociación han resultado ser muy interesantes para descubrir relaciones entre variables en grandes conjuntos de datos, no son un método 100% seguro, depende de la elección del analista determinar que reglas finales le serán de utilidad y cuáles no, por lo tanto, no es un método que yo usaría mucho fuera de temas relacionados al marketing y venta de productos, que es para lo que a mi parecer esta hecho este método, no son muy complicadas y con un buen conjunto de datos es una herramienta poderosa, el único inconveniente es que es un tanto tardado de elaborar, pero se pueden obtener resultados muy satisfactorios.

*Inserte una portada indicando el tema que aborda*

## 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	Nombre de la clase	Categorico	Demócrata, Republicano
bebés discapacitados	Voto para bebés discapacitados	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
costo compartido del proyecto de agua	Voto para costo compartido del proyecto de agua	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
adopción de la resolución presupuestaria	Voto para adopción de la resolución presupuestaria	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
congelación de honorarios médicos	Voto para congelación de honorarios médicos	Nominal	y=Si, n=No

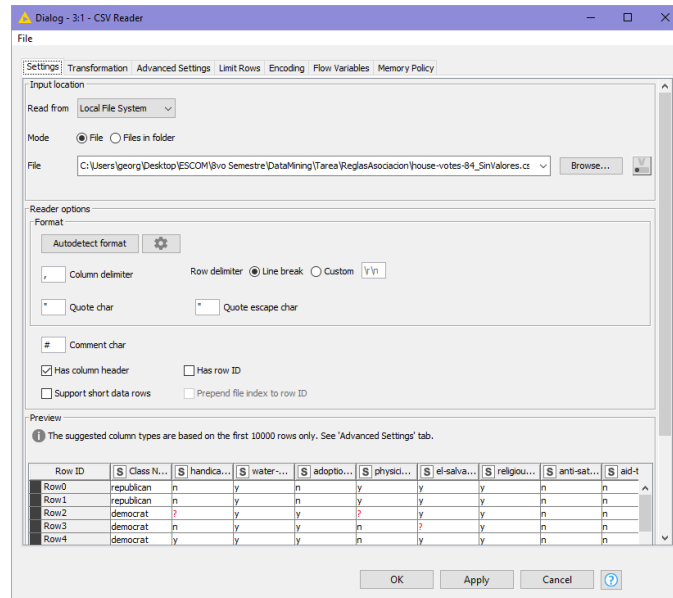
			?=El valor no es "sí" o "no"
ayuda el salvador	Voto para ayuda el salvador	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
grupos religiosos en las escuelas	Voto para grupos religiosos en las escuelas	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
prohibición de prueba anti-satélite	Voto para prohibición de prueba anti-satélite	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
ayuda a los contras nicaragüenses	Voto para ayuda a los contras nicaragüenses	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
misil mx	Voto para misil mx	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
inmigración	Voto para inmigración	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
reducción de la corporación synfuels	Voto para reducción de la corporación synfuels	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
gasto en educación	Voto para gasto en educación	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
superfondo derecho a demandar	Voto para superfondo derecho a demandar	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
crimen	Voto para crimen	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
exportaciones libres de impuestos	Voto para exportaciones libres de impuestos	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"



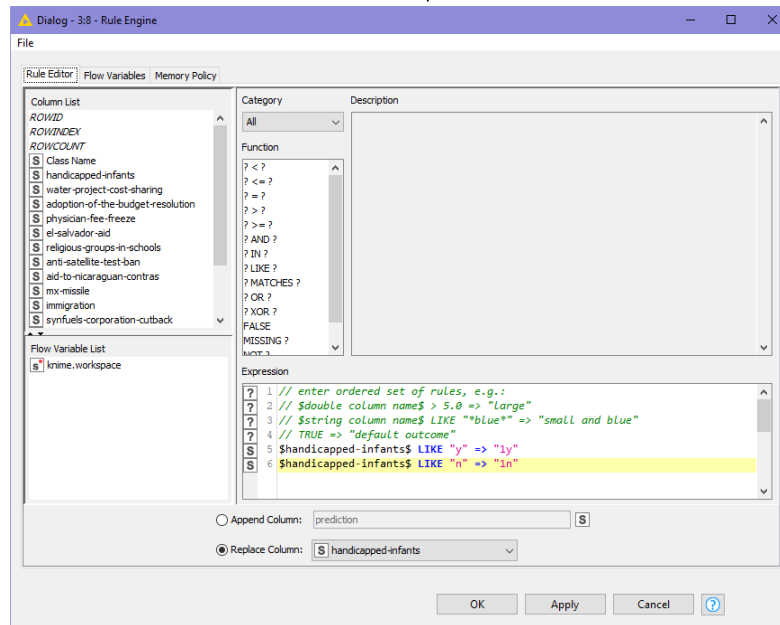
ley de administración de exportaciones de sudáfrica	ley de administración de exportaciones de sudáfrica	Nominal	y=Si, n=No ?=El valor no es "sí" o "no"
---	---	---------	---

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).

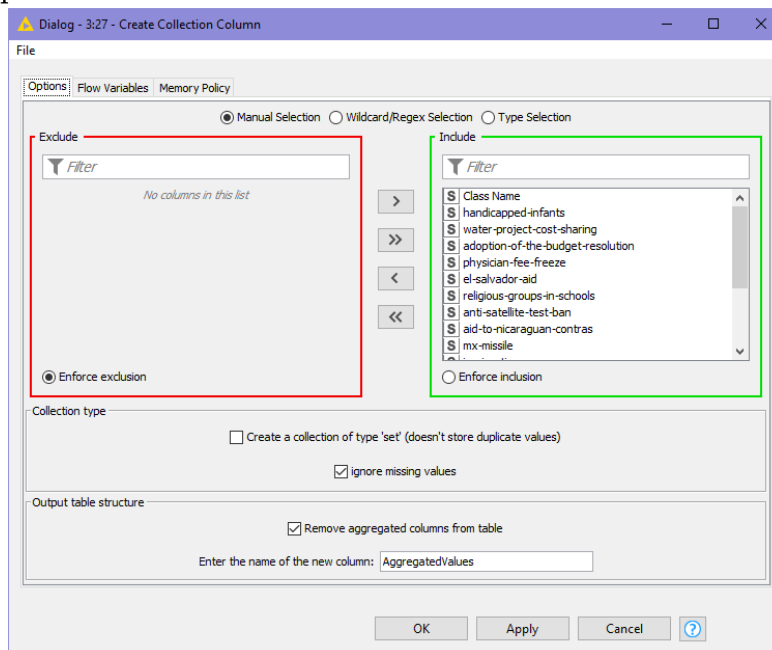
- Para empezar, cargamos el dataset en Knime, en forma de archivo CSV.



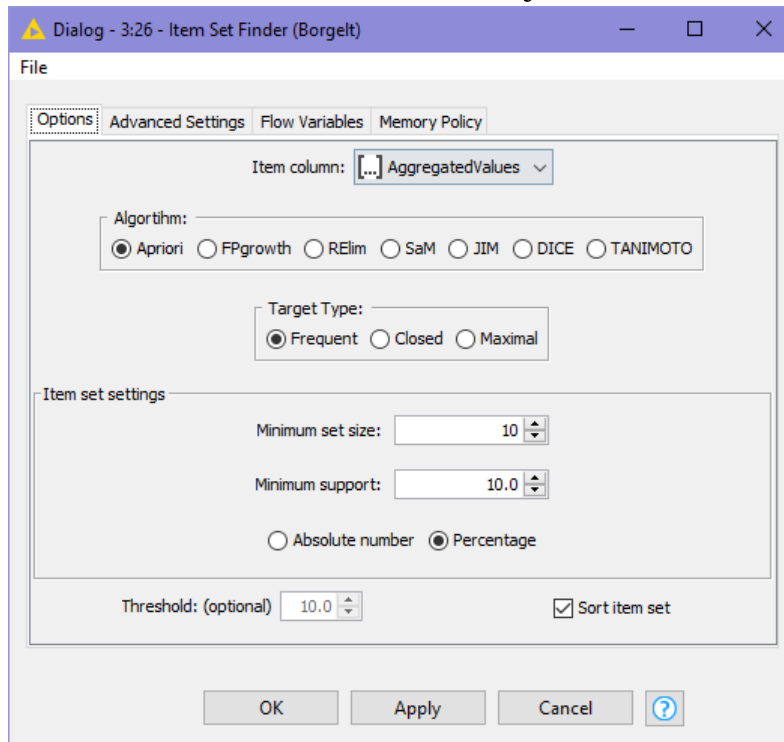
- Procedemos al tratamiento de los datos modificando los datos de las columnas de votación mediante los nodos: Rule Engine (Así para cada una de las 16 columnas).



- Después, creamos una colección con todas las columnas del dataset para terminar con una sola columna.



- 3) Elija las reglas de asociación que tengan como consecuencia las clases indicadas (Demócrata, Republicano) para identificar las características de cada partido político.
- Procedemos a la localización del conjunto de elementos.



- Procedemos a ejecutar el nodo Association Rule Learner.

Dialog - 3:24 - Association Rule Learner

File

Options Flow Variables Memory Policy

Itemset Mining

Column containing transactions: [...] ItemSet

Minimum support (0-1): 0.00

Underlying data structure: ARRAY

Output

Itemset type: CLOSED

Maximal itemset length: 2

Association Rules

☒ Output association rules

Minimum confidence: 0.3

OK Apply Cancel ?

- El anterior nodo nos va a generar la siguiente tabla:

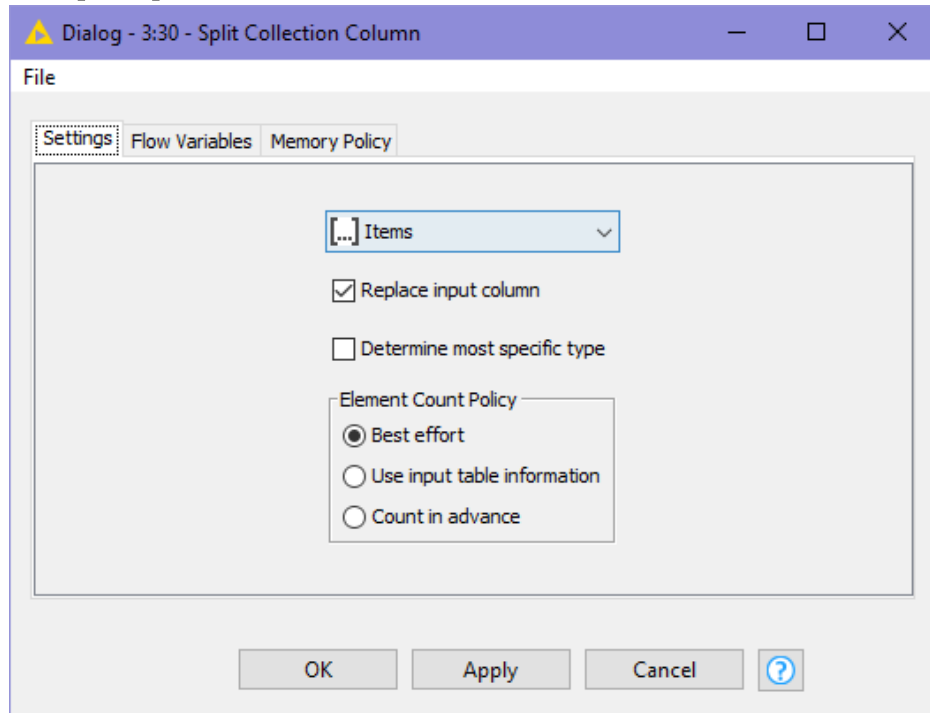
Frequent itemsets/Association rules - 3:24 - Association Rule Learner

File Edit Hilite Navigation View

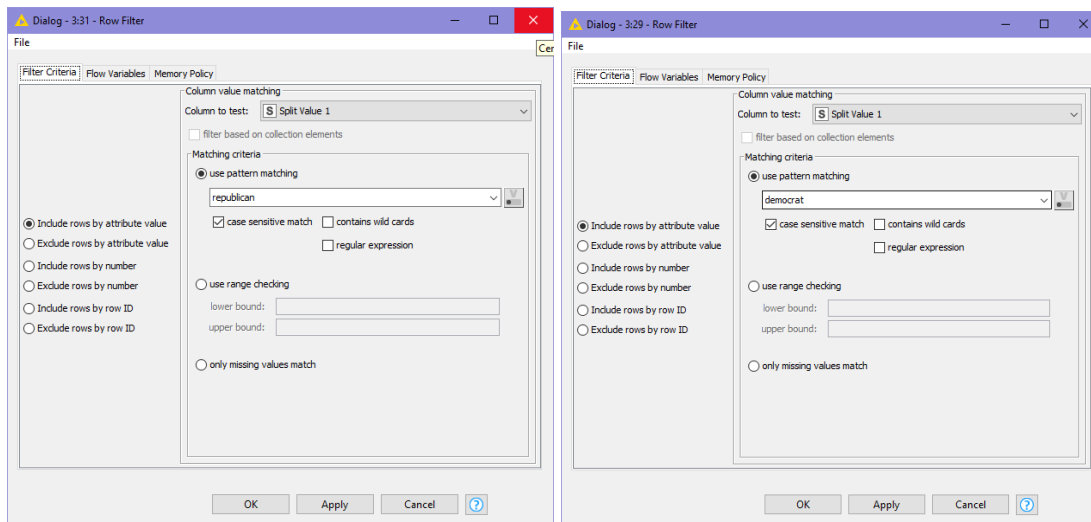
Table "default" - Rows: 400 Spec - Columns: 6 Properties Flow Variables

Row ID	[D] Support	[D] Confide...	[D] Lift	[S] Conseq...	[S] implies	[...] Items
rule375	0.334	0.726	1.67	8n	<---	[4y]
rule376	0.334	0.77	1.677	4y	<---	[14y]
rule377	0.334	0.728	1.677	14y	<---	[4y]
rule378	0.335	0.754	1.678	9n	<---	[6y]
rule379	0.335	0.746	1.678	6y	<---	[9n]
rule380	0.336	0.755	1.67	republican	<---	[6y]
rule381	0.336	0.742	1.67	6y	<---	[republican]
rule382	0.336	0.731	1.676	3n	<---	[4y]
rule383	0.336	0.77	1.676	4y	<---	[3n]
rule384	0.34	0.752	1.672	9n	<---	[republican]
rule385	0.34	0.756	1.672	republican	<---	[9n]
rule386	0.34	0.765	1.682	5y	<---	[6y]
rule387	0.34	0.748	1.682	6y	<---	[5y]
rule388	0.343	0.771	1.677	4y	<---	[6y]
rule389	0.343	0.746	1.677	6y	<---	[4y]
rule390	0.344	0.766	1.683	5y	<---	[9n]
rule391	0.344	0.757	1.683	9n	<---	[5y]
rule392	0.345	0.763	1.677	5y	<---	[republican]
rule393	0.345	0.758	1.677	republican	<---	[5y]
rule394	0.347	0.771	1.679	4y	<---	[9n]
rule395	0.347	0.755	1.679	9n	<---	[4y]
rule396	0.352	0.773	1.683	4y	<---	[5y]
rule397	0.352	0.766	1.683	5y	<---	[4y]
rule398	0.352	0.779	1.695	4y	<---	[republican]
rule399	0.352	0.766	1.695	republican	<---	[4y]

- Ahora vamos a separar la colección que previamente generamos, para poder filtrar los resultados.



- Procedemos a realizar un filtrado de las filas:



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

Filtered - 3:29 - Row Filter

File Edit Hilite Navigation View

Table "default" - Rows: 12 Spec - Columns: 6 Properties Flow Variables

Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	S Split Val...
rule84	0.145	0.425	2.408	15y	<---	democrat
rule106	0.17	0.499	2.312	1y	<---	democrat
rule170	0.229	0.674	2.363	13n	<---	democrat
rule176	0.232	0.682	2.353	6n	<---	democrat
rule198	0.245	0.719	2.349	14n	<---	democrat
rule222	0.253	0.743	2.359	12n	<---	democrat
rule232	0.255	0.749	2.37	9y	<---	democrat
rule253	0.261	0.768	2.362	7y	<---	democrat
rule263	0.266	0.783	2.364	3y	<---	democrat
rule268	0.27	0.794	2.365	5n	<---	democrat
rule271	0.27	0.794	2.363	8y	<---	democrat
rule272	0.274	0.804	2.363	4n	<---	democrat

Filtered - 3:31 - Row Filter

File Edit Hilite Navigation View

Table "default" - Rows: 13 Spec - Columns: 6 Properties Flow Variables

Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	S Split Val...
rule154	0.223	0.494	1.226	11n	<---	republican
rule237	0.257	0.568	1.638	1n	<---	republican
rule296	0.289	0.641	1.656	13y	<---	republican
rule308	0.295	0.652	1.669	12y	<---	republican
rule322	0.304	0.674	1.66	7n	<---	republican
rule351	0.324	0.716	1.673	15n	<---	republican
rule359	0.328	0.725	1.67	14y	<---	republican
rule360	0.328	0.726	1.669	8n	<---	republican
rule366	0.331	0.731	1.678	3n	<---	republican
rule381	0.336	0.742	1.67	6y	<---	republican
rule384	0.34	0.752	1.672	9n	<---	republican
rule392	0.345	0.763	1.677	5y	<---	republican
rule398	0.352	0.779	1.695	4y	<---	republican

Partido	Características (listado)
Demócrata	<ul style="list-style-type: none"> <li>• Si es demócrata existe 2.4 veces más probabilidad de que vote SÍ por exportaciones libres de derechos.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote SÍ por bebés discapacitados.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por superfondo-derecho-a-demandar.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por grupos-religiosos-en-escuelas.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por crimen.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por gasto en educación.</li> </ul>

	<ul style="list-style-type: none"> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote SÍ por misil mx.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote SÍ por prohibición-de-prueba-anti-satélite.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote SÍ por aprobación-del-presupuesto-resolución.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por ayuda del salvador.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote SÍ por ayuda a nicaragüenses contras.</li> <li>• Si es demócrata existe 2.3 veces más probabilidad de que vote NO por congelación de honorarios del médico.</li> </ul>
Republicano	<ul style="list-style-type: none"> <li>• Si es republicano existe 1.2 veces más probabilidad de que vote NO por recorte corporación synfuels.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por bebés discapacitados.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por superfondo derecho a demandar.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por gasto en educación.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por prohibición de prueba anti satélite.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por superfondo derecho a demandar.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por crimen.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por ayuda a nicaragüenses contras.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por aprobación del presupuesto resolución.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por grupos religiosos en escuelas.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote NO por misil mx.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por ayuda de el salvador.</li> <li>• Si es republicano existe 1.6 veces más probabilidad de que vote SÍ por congelación de honorarios del médico.</li> </ul>

### 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.

Considere el siguiente conjunto de datos:

Nombre:	Precios de calzado para mujer Una lista de 10,000 zapatos de mujer y los precios a los que se venden. <i>Women's Shoe Prices</i> <i>A list of 10,000 women's shoes and the prices at which they are sold.</i>
Objetivo:	Este conjunto de datos es una lista de 10,000 zapatos de mujer y su información de producto 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/datafiniti/womens-shoes-prices?select=Datafiniti_Womens_Shoes.csv">https://www.kaggle.com/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. <u>Esquema de datos</u> 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
--------	-------------	------	---------



ID	Id del item antecedente	int	0-1000
Nombre	Nombre del producto	string	Varios nombres de modelos de zapatos
Marca Acendente	Marca a la que pertenece el producto	string	Varias marcas de zapatos
Precio Total Acendente	Precio del producto	float	Suma del precio total del antecedente
ítem Recomendado	Item del producto recomendado	int	0-1000
Marca Recomendado	Marca del producto Recomendado	string	Varias marcas de zapatos
producto Recomendado	Nombre del producto Recomendado	string	Varios nombres de modelos de zapatos
precio Producto Recomendado	Precio del producto recomendado	float	Precio producto Recomendado
RuleRevenue	Recomendación del producto	float	4.96 - 1897.35

- 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,

#### **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.