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 |
| 2) | García Cruz Octavio Arturo |
| 3) | Sampayo Hernández Mauro |

Logotipo

Descripción generada automáticamenteImagen que contiene Logotipo

Descripción generada automáticamenteInstituto 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

3CV15

**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: Fi = 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

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

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

1. Desarrollo
   1. 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 |
| **8** | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| **9** | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| **10** | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| **Suma** | **6** | **7** | **5** | **6** | **6** | **5** | **4** |

* 1. Representación tabular

1. Primera Etapa

C1:

|  |  |
| --- | --- |
| **Itemset** | **Soporte** |
| 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 min\_sup = 2. min\_sup >= fi () = 2. Todos cumplen. Por lo tanto, el resultado es:

L1:

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

1. Segunda Etapa

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

|  |  |
| --- | --- |
| **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 | 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 |
| {Alexa, Reproductor de DVD} | 1 |
| {Consola de Videojuegos,Impresora} | 4 |
| {Consola de Videojuegos, Laptop } | 2 |
| {Consola de Videojuegos, Reproductor de DVD } | 1 |
| {Impresora, Laptop } | 4 |
| {Impresora, Reproductor de DVD } | 2 |
| { Laptop , Reproductor de DVD } | 3 |

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

L2:

|  |  |
| --- | --- |
| **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 |
| {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**

|  |  |
| --- | --- |
| **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 |
| {Consola de Videojuegos,Impresora} | 4 |
| {Consola de Videojuegos, Laptop } | 2 |
| {Impresora, Laptop } | 4 |
| {Impresora, Reproductor de DVD } | 2 |
| { Laptop , Reproductor de DVD } | 3 |

|  |  |
| --- | --- |
| **Itemset** | **Soporte** |
| {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 >= 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 >= 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:**

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

1. Evaluación de las reglas
2. Conclusiones

Logotipo

Descripción generada automáticamenteImagen que contiene Logotipo

Descripción generada automáticamenteInstituto 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

3CV15

**PARTE 2. Aprendizaje supervisado**

Considere el siguiente conjunto de datos:

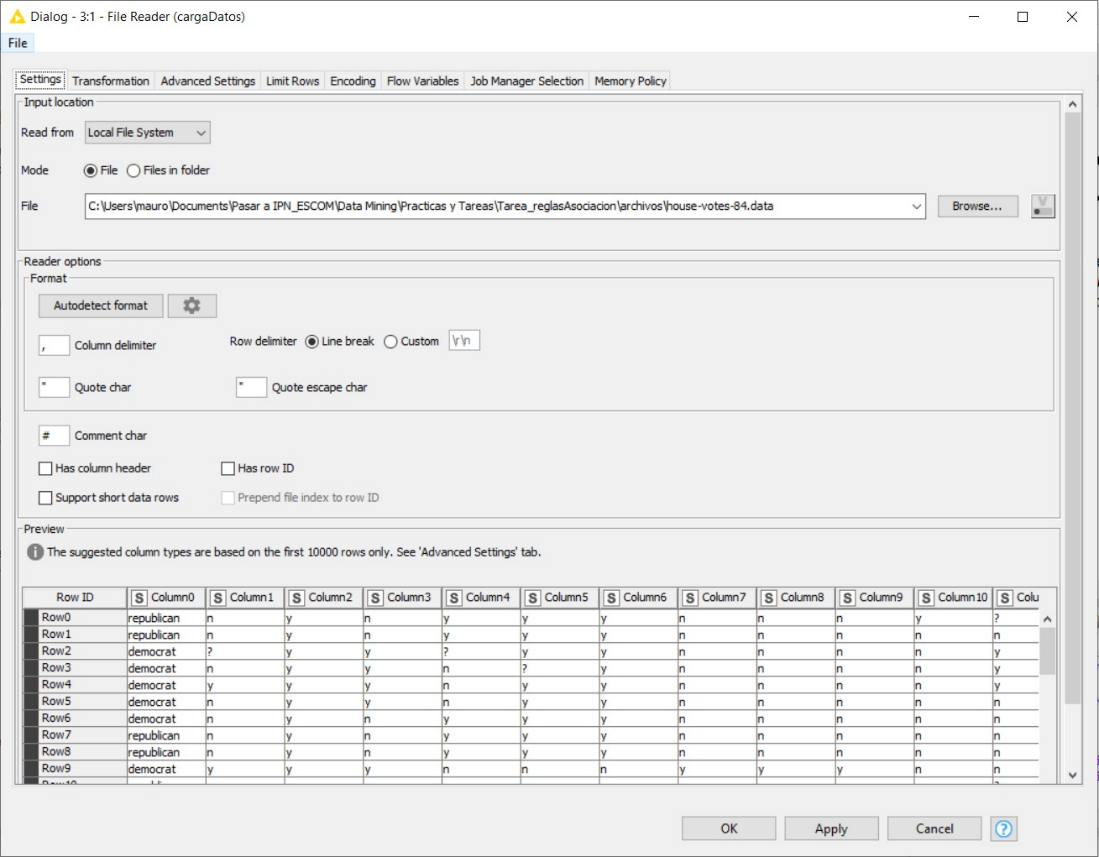
|  |  |
| --- | --- |
| Nombre: | Conjunto de datos de registros de votación del Congreso  (*Congressional Voting Records Data Set*) |
| Objetivo: | Este conjunto de datos incluye los votos de cada uno de los congresistas de la Cámara de Representantes de los Estados Unidos sobre los 16 votos clave identificados 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: | <https://archive.ics.uci.edu/ml/datasets/congressional+voting+records> |
| Clases: | Demócrata, Republicano |

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

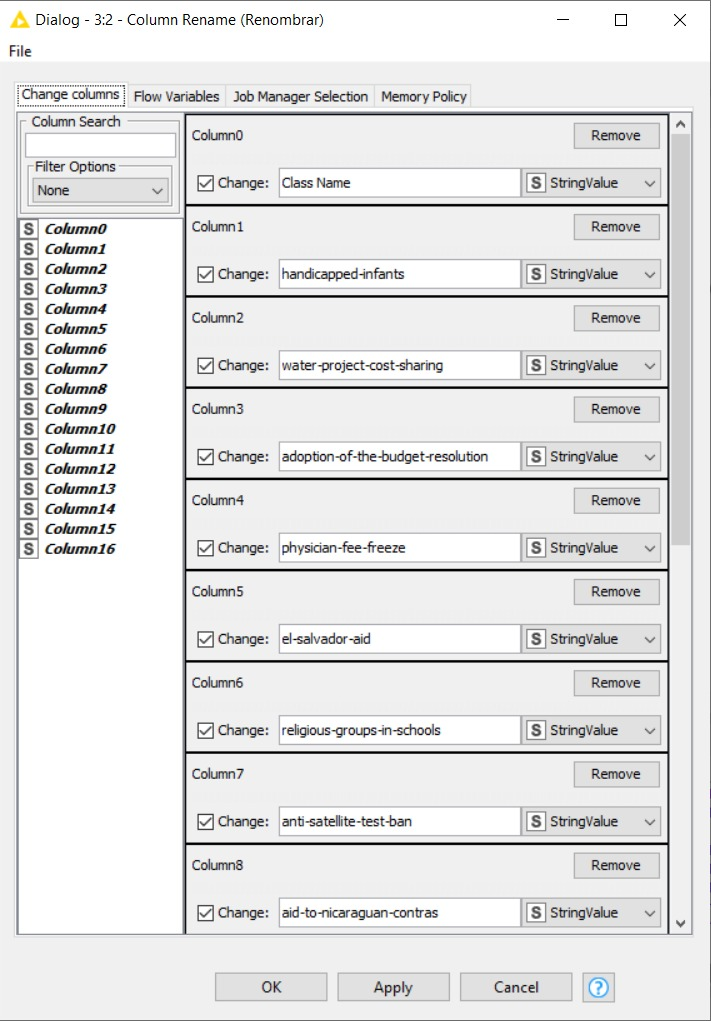
|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre** | **Significado** | **Tipo** | **Dominio** |
| **Nombre de la clase** | Describe a que partido político pertenece | Categórico | Demócrata, Republicano |
| **Niños discapacitados** | Apoyo a Niños Discapacitados | Categórico | Yes, No |
| **Costos 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 |
| **Prohibición de pruebas anti-satélite** | Esta a favor de la Prohibición de pruebas anti-satélite | Categórico | Yes, No |
| **Ayuda-a-los-contras-nicaraguenses** | Propone ayuda a las gentes Nicaraguenses | Categórico | Yes, No |
| **mx-misil** | A favor o en contra de mx-misil | Categórico | Yes, No |
| **inmigración** | Apoya o esta en contra de la inmigración hacia Estados Unidos | Categórico | Yes, No |
| **synfuels-corporation-cutback** | Apoya las synfuels-corporation-cutback | Categórico | Yes, No |
| **gasto en educación** | Prone reducir el gasto en la educacion | Categórico | Yes, No |
| **superfund-right-to-sue** | Apoya las superfund-right-to-sue | Categórico | Yes, No |
| **crimen** | Propuestas para la reducción del crimen | Categórico | Yes, No |
| **exportaciones libres de impuestos** | Apoya las exportaciones libres de impuestos | Categórico | Yes, No |
| **Ley de administración de exportaciones-Sudáfrica** | A favor o en contra de la Ley de administración de exportaciones-Sudáfrica | Categórico | Yes, No |

1. 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



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



Una vez completado el tratamiento anterior procedemos a conectar nuestro colum 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**)**

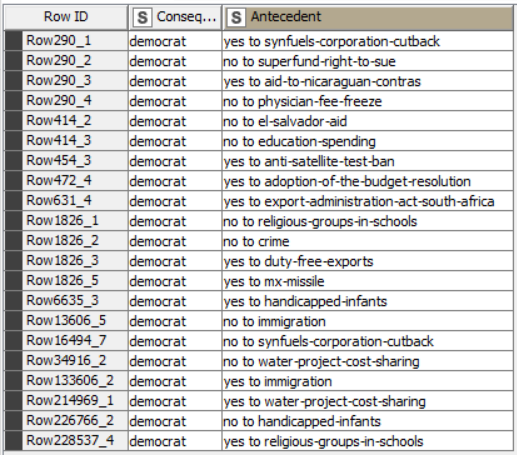
# Tabla de Salida

knio**.**output\_tables**[**0**]** **=** knio**.**Table**.**from\_pandas**(**votos**)**

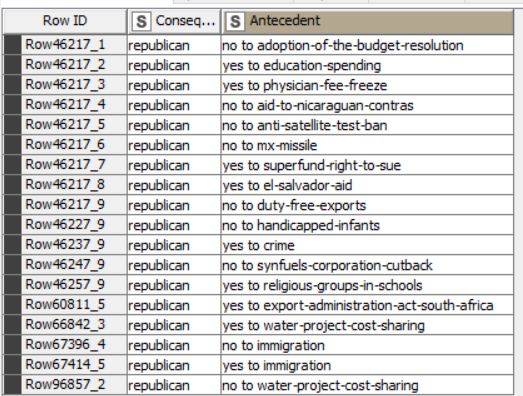
1. Elija las reglas de asociación que tengan como consecuente las clases indicadas (Demócrata, Republicano) para identificar las características de cada partido político.

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

**Demócratas**



**Republicanos**



1. Describa las características generadas por las reglas de asociación que distinguen cada partido.

|  |  |
| --- | --- |
| **Partido** | **Características (listado)** |
| Demócrata | * **EN CONTRA** de la ley *superfund right to sue.* * **EN CONTRA** de la *congelación de honorarios médicos.* * **EN CONTRA** de la *ayuda a El Salvador.* * **EN CONTRA** del *gasto en la educación.* * **EN CONTRA** de *tomar medidas para combatir el crimen.* * **A FAVOR** de la *ayuda a los contras nicaragüenses.* * **A FAVOR** de la *prohibición de pruebas antisatélite.* * **A FAVOR** de la *adopción de la resolución presupuestaria.* * **A FAVOR** de *la ley de administración de exportación a Sudáfrica. \** * **A FAVOR** de las *exportaciones libres de impuestos.* * **A FAVOR** de los *misiles-mx.* * Existe una **DIVISIÓN MARCADA** aparente, entre los simpatizantes de este partido, que están **a favor** y **en contra** de *la inmigración, la reducción de ingresos destinados a la Synthetic Fuels Corporation, el proyecto de costos compartidos de agua, el apoyo a los niños discapacitados* y *la presencia de grupos religiosos en escuelas.* |
| Republicano | * **EN CONTRA** de la *adopción de la resolución presupuestaria.* * **EN CONTRA** de la *ayuda a los contras nicaragüenses.* * **EN CONTRA** de la *prohibición de pruebas antisatélite.* * **EN CONTRA** de los *misiles-mx.* * **EN CONTRA** de las *exportaciones libres de impuestos.* * **EN CONTRA** del *apoyo a los niños discapacitados.* * **EN CONTRA** de *la reducción de ingresos destinados a la Synthetic Fuels Corporation.* * **A FAVOR** del *gasto en la educación.* * **A FAVOR** de la *congelación de honorarios médicos.* * **A FAVOR** de la ley *superfund right to sue.* * **A FAVOR** de la *ayuda a El Salvador.* * **A FAVOR** de *tomar medidas para combatir el crimen.* * **A FAVOR** de *la presencia de grupos religiosos en escuelas.* * **A FAVOR** de *la ley de administración de exportación a Sudáfrica. \** * Existe una **DIVISIÓN MARCADA** aparente, entre los simpatizantes de este partido, que están **a favor** y **en contra** de *la inmigración* y *el proyecto de costos compartidos de agua.* |

Agregue las reglas de asociación que generó tales características

Logotipo

Descripción generada automáticamenteImagen que contiene Logotipo

Descripción generada automáticamenteInstituto 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

3CV15

**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 à 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: | <https://www.kaggle.com/datasets/datafiniti/womens-shoes-prices?select=Datafiniti_Womens_Shoes.csv>  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.  <https://developer.datafiniti.co/docs/product-data-schema> |

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 fuente de información | Texto | 2 lips too  Adidas  Adidas outdoor  Altra  .  .  .  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 utiliza para combinar datos sin procesar de fuentes individualesen el registro maestro de Datafiniti | Palabra clave | 000000000000 - 999999999999 |
| 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 |

1. 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.

1. 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>

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 on 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.