

INSTITUTO POLITÉCNICO NACIONAL  
ESCUELA SUPERIOR DE CÓMPUTO  
Unidad de Aprendizaje: Minería de datos  
Ciclo escolar: 2023-1

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

Grupo: 3CV11

Equipo: \_\_\_\_\_

Nombre de los integrantes del equipo:

1) \_\_\_\_\_

2) \_\_\_\_\_

3) \_\_\_\_\_

*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:  $F_i = 2$ ,  $\text{min.sup} = 2$ ,  $\text{min.conf} = 75\%$  (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
- 2) Desarrollo
  - a. Listado de transacciones
  - b. Representación tabular
- 3) Primera Etapa
- 4) Segunda Etapa
- 5) Evaluación de las reglas
- 6) Conclusiones

*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

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

Partido	Características (listado)
Demócrata	– –
Republicano	– –

Agregue las reglas de asociación que generó tales características

*Inserte una portada indicando el tema que aborda*

### **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

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