

Minería de datos

Tema 5: Análisis de asociación



Agenda:

- **Análisis de asociación**
- **Conceptos básicos**
- **Aprendizaje de reglas de asociación**
- **Algoritmo A priori**
- **Evaluación de reglas de asociación**
- **Ejemplos de aplicación**

- **Dada una colección de registros o instancias**

Donde cada registro contiene un conjunto de atributos o items, no hay salida definida

- **Encontrar combinaciones o asociaciones de items (atributos) que ocurren frecuentemente.**

- **Objetivo: descubrir patrones que describen características fuertemente asociadas en los datos.**

Útil para descubrir relaciones interesantes ocultas en los datos

Estas relaciones expresan patrones de comportamiento en función de la aparición conjunta de valores de dos o más atributos

El análisis de asociación es aplicable a muchos dominios, por ejemplo:

- **Búsqueda de patrones en páginas Web:**

Determinar cuáles son los itinerarios más seguidos por los visitantes de un sitio Web.

- **Análisis de peticiones de servicios médicos:**

Encontrar aquellas pruebas o exámenes médicos que frecuentemente se realizan juntas.

- **Análisis de ventas de productos:**

Determinar la frecuencia de clientes que al comprar un producto A, seis meses después compran un producto B.

- **Ciencias de la tierra:**

Determinar las conexiones más interesantes entre cielo, tierra y procesos atmosféricos.

- **Análisis de la cesta de compra:**

Encontrar los productos que se compran juntos más frecuentemente

- ¿Cómo son las reglas de asociación?

Ejemplo 1: análisis de la cesta de compra

ID	Items
1	{Pan, leche}
2	{Pan, servilletas, cerveza, huevos}
3	{leche, servilletas, cerveza, agua}
4	{Pan, leche, servilletas, cerveza}
5	{Pan, leche, servilletas, agua}

Analizar estos datos puede ser de gran utilidad, ya que a partir de ellos se puede aprender el comportamiento de compras de los clientes

Esta información puede ser utilizada para la toma de decisiones sobre promociones de mercadeo, gestión de inventarios y manejo de las relaciones con los clientes



Cada fila = transacción = conjunto de *items* comprados por un cliente dado

Una posible regla sería:

Si pan y servilletas entonces cerveza

Ejemplo 2: Análisis de petición de servicios médicos en una institución

Paciente	Perfil 20	Orina	Heces	Colesterol	VIH	Glicemia
1	X	X	0	0	0	0
2	X	0	X	X	X	0
3	0	X	X	X	0	X
4	X	X	X	X	0	0
5	X	X	X	0	0	X

En este caso, identificar las pruebas médicas que frecuentemente se realizan juntas puede ayudar a mejorar la gestión de la institución.

Una posible regla sería:

Si Orina entonces Perfil 20

- Términos básicos:

Los datos de la cesta de compras (y en general, los datos de transacciones), se pueden representar como una tabla binaria:

	IDt	Pan	Leche	Servilleta	Cerveza	Huevos	Agua
t_1	1	1	1	0	0	0	0
t_2	2	1	0	1	1	1	0
t_3	3	0	1	1	1	0	1
t_4	4	1	1	1	1	0	0
t_5	5	1	1	1	0	0	1

Un ítem = variable binaria

Se tiene entonces

$I = \{i_1, i_2, i_3, \dots, i_d\}$ = Conjunto de ítems

$T = \{t_1, t_2, t_3, \dots, t_N\}$ = Conjunto de transacciones

Cada transacción t_i contiene un subconjunto de ítems de I

👉 **Conjunto de ítems**: Conjunto de uno o más ítems.

Ejemplo: {leche, pan, servilletas}

👉 **Conjunto de k ítems**: Conjunto que contiene exactamente k ítems

👉 **Soporte de un conjunto de ítems**: Número de transacciones que contienen ese conjunto de ítems (= frecuencia de ocurrencia de un conjunto de ítems)

Si X es un conjunto de ítems entonces su soporte será

$$\sigma(X) = | \{ t_i / X \subseteq t_i, t_i \in T \} |$$

Ejemplo: $\sigma(\{\text{cerveza, servilletas, leche}\}) = | \{ t_3, t_4 \} | = 2$

👉 **Conjunto de ítems frecuente**: Conjunto de ítems cuyo soporte es igual o mayor que un umbral SOP_{Min}

- ¿Cómo evaluar una regla de asociación?

Para la evaluación de las reglas de asociación se pueden utilizar las siguientes métricas:

Sea la regla de la forma $X \longrightarrow Y$, donde X e Y son conjuntos de ítems

👉 **Soporte**: fracción de transacciones que contienen a X y a Y .
Determina cuán frecuentemente una regla es aplicable a un conjunto de datos.

$$\text{soporte } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

👉 **Confianza**: medida de cuán frecuentemente los ítems en Y aparecen en transacciones que contienen a X . Mide la confiabilidad de una regla.

$$\text{confianza } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

- El soporte de una regla es una medida muy importante. Reglas con bajo soporte no son significativas.
- El soporte proporciona un estimado de la probabilidad $P(X \cup Y)$
- A más alta la confianza, será más probable que Y esté presente en transacciones que contienen a X.
- La confianza también proporciona un estimado de la probabilidad condicional de Y dado X, $P(Y|X)$

Ejemplo:

IDt	Pan	Leche	Servilleta	Cerveza	Huevos	Agua
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

Si se tiene la regla $R: \{\text{leche, servilletas}\} \longrightarrow \{\text{cerveza}\}$

- *Soporte de la regla:*

$$\text{soporte}(R) = \frac{\sigma(\{\text{cerveza, servilletas, leche}\})}{5} = \frac{2}{5} = 0.40$$

- *Confianza de la regla:*

$$\text{confianza}(R) = \frac{\sigma(\{\text{cerveza, servilletas, leche}\})}{\sigma(\{\text{servilletas, leche}\})} = \frac{2}{3} = 0.67$$

¿Cómo se interpreta? ➡ 67% de las veces que se compra leche y servilletas también se compra cerveza y esto ocurre en el 40% de los casos.

Aprendizaje de reglas de asociación

- ¿Cómo extraer reglas de asociación a partir de datos?

La tarea de minería de reglas de asociación se basa en:

Dado un conjunto de transacciones T , encontrar todas las reglas que tengan:

$$\begin{aligned} \text{Soporte} &\geq \text{Sop}_{\text{Min}} \\ \text{Confianza} &\geq \text{Conf}_{\text{Min}} \end{aligned}$$

Umbral para el soporte y la confianza de las reglas. Definidas por el usuario o experto del dominio.

Un posible enfoque:

- Determinar todas las posibles reglas
- Calcular el soporte y la confianza para cada regla
- Eliminar las reglas que no satisfacen los umbrales para el soporte y la confianza

➡ ¡Muy ineficiente!

Aprendizaje de reglas de asociación

La estrategia más común adoptada por muchos algoritmos de minería de reglas de asociación

➡ **Descomponer el problema en dos subtareas**

1.- Generación de los conjuntos de ítems frecuentes:

- **Encontrar todos los conjuntos de ítems que satisfacen el umbral Sop_{Min}**

El soporte de una regla depende sólo del soporte de su correspondiente conjunto de ítems $(X \cup Y)$.

2.- Generación de las reglas:

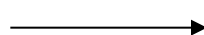
- **Encontrar todas las reglas de alta confianza (que cumplan con el umbral para la confianza $Conf_{Min}$), a partir de los conjuntos de ítems frecuentes encontrados en el paso previo. Estas reglas se llaman **reglas fuertes (strong rules)**.**

Aprendizaje de reglas de asociación

Ejemplo:

IDt	Pan	Leche	Servilleta	Cerveza	Huevos	Agua
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

Soporte mínimo: 0.40



Soporte de los CI = $\sigma = 2$ (0.40 x 5)

Confianza mínima: 0.60

➡ **Algunos conjuntos de ítems frecuente con soporte mayor o igual a 0.40:**

{pan, leche}

{pan, servilleta}

{pan, cerveza}

{leche, cerveza}

{cerveza, servilletas, leche}

{pan, servilletas, cerveza}

Aprendizaje de reglas de asociación

Del conjunto: {cerveza, servilletas, leche}

Se pueden derivar las siguientes reglas:

Confianza

{cerveza} \longrightarrow {servilletas, leche}	$(2/3) = 0.66$
{servilletas} \longrightarrow {cerveza, leche}	$(2/4) = 0.50$
{leche} \longrightarrow {cerveza, servilletas}	$(2/4) = 0.50$
{cerveza, servilletas} \longrightarrow {leche}	$(2/3) = 0.66$
{cerveza, leche} \longrightarrow {servilletas}	$(2/2) = 1.00$
{servilletas, leche} \longrightarrow {cerveza}	$(2/3) = 0.66$

Las reglas seleccionadas serían:

- R1: Si {cerveza} entonces {servilletas, leche}
- R2: Si {cerveza, servilletas} entonces {leche}
- R3: Si {cerveza, leche} entonces {servilletas}
- R4: Si {servilletas, leche} entonces {cerveza}

Aprendizaje de reglas de asociación

- *¿Cómo determinar un modelo de asociación?*

Algunas técnicas:

- **Apriori**
- **FP-Growth**
- **Eclat**
- **TERTIUS**
- **Algoritmos genéticos**
- **Métodos para descubrir reglas de asociación secuenciales, entre otros**

Primer algoritmo de minería de reglas de asociación en utilizar la búsqueda basada en el soporte de los conjuntos de ítems.

A) Generación de los conjuntos de ítems frecuentes

Ejemplo: Sea $I = \{a, b, c, d\}$

Se deben tomar en cuenta todas las posibles combinaciones de ítems



{ }					
{a}	{b}	{c}	{d}		
{a, b}	{a, c}	{a, d}	{b, c}	{b, d}	{c, d}
{a, b, c}	{a, b, d}	{a, c, d}	{b, c, d}		
{a, b, c, d}					

¿Cómo reducir la complejidad de la búsqueda?



Reducir el número de posibles ítems candidatos

Principio Apriori: Utilizar el soporte para reducir el número de conjuntos de ítems (CI) explorados durante la generación de los conjuntos de ítems frecuentes

Entonces,

a) Si un conjunto de ítems es frecuente entonces todos sus subconjuntos también serán frecuentes.

Ejemplo: si $\{b, c, d\}$ es un CI frecuente (cumple con el umbral para el soporte)

entonces $\{b\}$, $\{c\}$, $\{d\}$, $\{b, c\}$, $\{b, d\}$, $\{c, d\}$ también serán CI frecuentes

b) Si un conjunto de ítems es infrecuente entonces todos sus superconjuntos (aquellos que lo contienen) también serán infrecuentes.

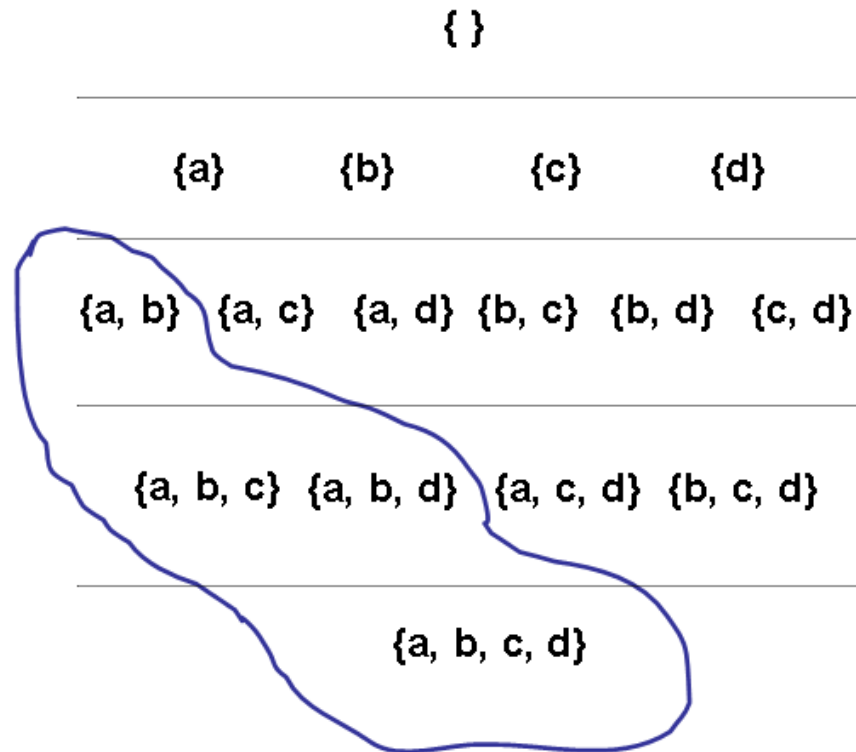
Ejemplo: si $\{a, b\}$ es un CI infrecuente (no cumple con el umbral para el soporte)

entonces $\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, c, d\}$ también serán CI infrecuentes

El principio b) permite mejorar la búsqueda en el espacio de CI candidatos

Este principio se utiliza para evaluar cada posible nodo (CI) del árbol de búsqueda:

Si {a, b} es un CI infrecuente, no tiene sentido seguir la búsqueda por esta rama



Ejemplo:

IDt	Pan	Leche	Servilleta	Cerveza	Huevos	Agua
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

1. Inicialmente cada ítem se considera como un CI de 1 ítem (1-itemset)

$Sop_{Min} = 60\%$ → El soporte que deben cumplir los CI es:

$$N \times 0.60 = 5 \times 0.60 = 3$$

1-itemset	Soporte
{cerveza}	3
{pan}	4
{agua}	2
{servilleta}	4
{leche}	3
{huevos}	1

Se descartan, ya que no cumplen el soporte mínimo

2. En la siguiente iteración, los CI con 2 ítems (2-itemset) se generan utilizando sólo los conjuntos de ítems frecuentes de 1 ítem, debido al principio Apriori

2-itemset	Soporte
{cerveza, pan}	2
{cerveza, servilleta}	3
{cerveza, leche}	2
{pan, servilleta}	3
{pan, leche}	3
{servilletas, leche}	3

Se descartan, ya que no cumplen el soporte mínimo

3. En la siguiente iteración, los CI con 3 ítems (3-itemset) se generan utilizando sólo los conjuntos de ítems frecuentes de 2 ítems.

3-itemset	Soporte
{cerveza, servilleta, pan}	2
{cerveza, servilleta, leche}	2
{pan, servilleta, leche}	2

Se descartan, ya que no cumplen el soporte mínimo



Conjuntos de ítems frecuentes, derivados a partir de los datos:

{cerveza, servilleta}

{cerveza, leche}

{pan, leche}

{servilletas, leche}

B) Generación de las reglas de asociación

¿Cómo extraer reglas de asociación de manera eficiente a partir de los conjuntos de ítems frecuentes?

- Las reglas que son generadas a partir de CI frecuentes satisfacen el umbral del soporte.
- Cada CI de k ítems puede generar $(2^k - 2)$ reglas.

Ejemplo: del conjunto {cerveza, servilletas, leche}

Se pueden derivar las siguientes reglas:

{cerveza} \longrightarrow {servilletas, leche}
 {servilletas} \longrightarrow {cerveza, leche}
 {leche} \longrightarrow {cerveza, servilletas}
 {cerveza, servilletas} \longrightarrow {leche}
 {cerveza, leche} \longrightarrow {servilletas}
 {servilletas, leche} \longrightarrow {cerveza}

- Además, calcular la confianza de una regla de asociación no requiere búsquedas adicionales en el conjunto de datos.

Ejemplo:

Para la regla {cerveza, servilletas} \longrightarrow {leche}

Generada a partir del CI frecuente {leche, servilletas, cerveza}

La confianza será
$$\frac{\sigma(\{\text{cerveza, servilletas, leche}\})}{\sigma(\{\text{cerveza, servilletas}\})}$$

Es un CI frecuente debido al principio Apriori

El soporte de ambos ya fue calculado durante la generación de los CI frecuentes

Para explorar el espacio de posibles reglas se puede utilizar el siguiente teorema:

Sea Y un CI frecuente, entonces:

Si una regla $X \longrightarrow Y - X$ no satisface el umbral de la confianza entonces cualquier regla $X' \longrightarrow Y - X'$ (donde $X' \subseteq X$), tampoco lo hará.

Ejemplo:

$Y = \{\text{cerveza, servilletas, leche, pan}\}$

Si la regla: $\{\text{cerveza, servilletas, leche}\} \longrightarrow \{\text{pan}\}$

No satisface el umbral para la confianza, entonces tampoco lo hará la regla:

$\{\text{cerveza, servilletas}\} \longrightarrow \{\text{pan, leche}\}$

Donde: $X = \{\text{cerveza, servilletas, leche}\}$

$X' = \{\text{cerveza, servilletas}\}$

Demostración:

La confianza de las reglas es:

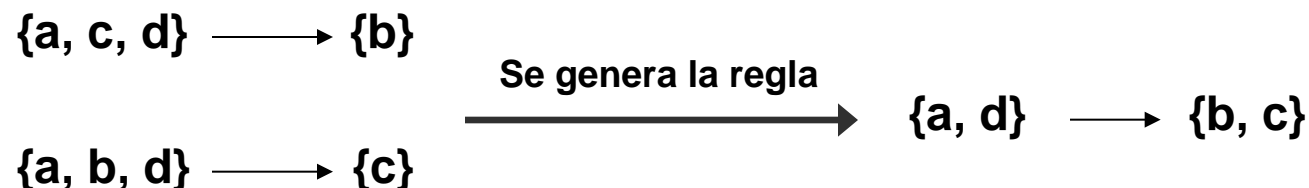
$$\left. \begin{array}{l} \frac{\sigma(\{X\} \cup \{Y - X\})}{\sigma(\{X\})} = \frac{\sigma(\{Y\})}{\sigma(\{X\})} \\ \frac{\sigma(\{X'\} \cup \{Y - X'\})}{\sigma(\{X'\})} = \frac{\sigma(\{Y\})}{\sigma(\{X'\})} \end{array} \right\} \begin{array}{l} \text{Como } X' \subseteq X \text{ entonces} \\ \sigma(\{X'\}) \geq \sigma(\{X\}) \\ \downarrow \\ \frac{\sigma(\{Y\})}{\sigma(\{X\})} \geq \frac{\sigma(\{Y\})}{\sigma(\{X'\})} \end{array}$$

La confianza de la regla $X' \rightarrow Y - X'$ no puede ser mayor que la confianza de la regla $X \rightarrow Y - X$ (donde $X' \subseteq X$).

- En el algoritmo Apriori se utiliza un enfoque por niveles para generar reglas de asociación, donde cada nivel corresponde al número de ítems en el consecuente de una regla.
- Inicialmente se generan todas las reglas de alta confianza con un ítem como consecuente. Estas reglas son entonces utilizadas para generar las reglas con dos ítems en el consecuente y así sucesivamente

Ejemplo:

De las reglas



Entonces, el árbol de búsqueda para el CI $\{a, b, c, d\}$ será

$$\{a, b, c, d\} \longrightarrow \emptyset$$

$$\{b, c, d\} \longrightarrow \{a\}$$

$$\{a, c, d\} \longrightarrow \{b\}$$

$$\{a, b, d\} \longrightarrow \{c\}$$

$$\{a, b, c\} \longrightarrow \{d\}$$

$$\{c, d\} \longrightarrow \{a, b\}$$

$$\{b, c\} \longrightarrow \{a, d\}$$

$$\{a, c\} \longrightarrow \{b, d\}$$

$$\{b, d\} \longrightarrow \{a, c\}$$

$$\{a, d\} \longrightarrow \{b, c\}$$

$$\{a, b\} \longrightarrow \{c, d\}$$

$$\{d\} \longrightarrow \{a, b, c\}$$

$$\{c\} \longrightarrow \{a, b, d\}$$

$$\{b\} \longrightarrow \{a, c, d\}$$

$$\{a\} \longrightarrow \{b, c, d\}$$

Puede ser eliminado

Si la confianza de $\{b, c, d\} \longrightarrow \{a\}$ no supera el umbral entonces
 todas las reglas que tengan al ítem "a" en el consecuente tampoco lo harán

Ejemplo: {cerveza, servilleta, leche} es CI frecuente

¿Cuántas reglas se pueden generar si la confianza es del 70%?

1er. Nivel:

Confianza:

{cerveza, servilletas} → {leche}

$(2/3) = 0.66$

{cerveza, leche} → {servilletas}

$(2/2) = 1.00$ ←

{servilletas, leche} → {cerveza}

$(2/3) = 0.66$

Se considera sólo esta regla en el siguiente nivel

2do. Nivel:

Confianza:

{cerveza} → {servilletas, leche}

$(2/3) = 0.66$

{leche} → {cerveza, servilletas}

$(2/3) = 0.66$

➡ La única regla sería {cerveza, leche} → {servilletas}

Este procedimiento se repite para cada CI frecuente

- El algoritmo, por cada regla generada, indica cual es el soporte y la confianza.
- Esta información puede ser utilizada para “interpretar la regla”
- En el ejemplo de análisis de petición de servicios de laboratorio, si el soporte y la confianza de la regla es del 60% y 70%, respectivamente, se podría interpretar de la siguiente forma:

Si **Orina** entonces **Perfil 20** (soporte =60%, confianza = 70%)



70% de las veces que se pide un análisis de orina también se solicita uno de sangre, y esto ocurre en 6 de cada 10 pacientes

Evaluación de reglas de asociación

Sea la regla: $X \rightarrow Y$ (soporte, confianza)

Donde,

a) **Soporte**: fracción de transacciones que contienen a X y a Y.

$$\text{soporte } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

b) **Confianza**: mide cuán frecuentemente los ítems en Y aparecen en transacciones que contienen a X.

$$\text{confianza } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

Evaluación de reglas de asociación

c) Medidas de correlación (toman en cuenta la dependencia estadística)

➡ $X \rightarrow Y$ (soporte, confianza, correlación)

Tomar en cuenta que:

$$\text{confianza}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} = P(Y|X) = \frac{P(X, Y)}{P(X)}$$

Se define la medida de sustentación o *lift* como:

$$\text{lift}(X \rightarrow Y) = \frac{\text{confianza}(X \rightarrow Y)}{\text{soporte}(Y)} = \frac{P(Y|X)}{P(Y)} = \frac{P(X \cup Y)}{P(X) P(Y)}$$

Evaluación de reglas de asociación

Esta medida evalúa el grado de dependencia entre los ítems o términos de una regla.

La evaluación de una regla utilizando *lift* se realiza de la siguiente manera:

$$\text{Si } \textit{lift}(X \rightarrow Y) = \begin{cases} 1, & \text{entonces la ocurrencia de Y es independiente de la} \\ & \text{ocurrencia de X, y viceversa (eventos independientes,} \\ & \text{no correlación entre ellos)} \\ >1, & \text{la ocurrencia de X implica la ocurrencia de Y (están} \\ & \text{positivamente correlacionados)} \\ < 1, & \text{entonces la ocurrencia de X influye en la} \\ & \text{probabilidad de no ocurrencia de Y (están} \\ & \text{negativamente correlacionados, la ocurrencia de uno} \\ & \text{implica la ausencia del otro)} \end{cases}$$

Se pueden utilizar otras medidas, tales como: la convicción, el interés, laplace, coseno, kappa, entre otros.

- La minería de datos aplicada al descubrimiento de patrones de supervivencia en mujeres con cáncer invasivo de cuello uterino

(Timarán, R. y Yépez, M. (2012). Revista Universidad y Salud. 14(2):117 – 129. ISSN: 0124-7107)

Se aplica la minería de datos para la extracción de reglas de asociación, utilizando el algoritmo Apriori (con la herramienta WEKA), para determinar patrones (factores socioeconómicos y clínico) de supervivencia en mujeres con cáncer invasivo de cuello uterino. Se utiliza también el algoritmo C4.5 para estimar tiempo de supervivencia.

Para los casos de mujeres que sobreviven al cáncer invasivo de cuello uterino, con una confianza mínima del 80% y un soporte mínimo del 20%, algunas de las relaciones encontradas fueron:

El 100% de las mujeres que viven en casa o apartamento y su ocupación es hogar, pertenece al régimen de salud subsidiado. El 30.2% de todas las mujeres que sobrevive tiene estas características.

El 100% de las mujeres cuya escolaridad es primaria y no tienen ninguna discapacidad, pertenece al régimen de salud subsidiado. El 30% de todas las mujeres que sobrevive tiene estas características.

➔ **Estudio descriptivo del nivel de interacción de los estudiantes de postgrado en la plataforma Moodle de la Facultad de Ciencias**

(Luis Arredondo)

PROBLEMA:

☞ **¿Cómo los estudiantes utilizan estas herramientas?**

☞ **¿Este uso puede dar evidencias de un perfil de los estudiantes?**

Objetivo: Determinar diferentes comportamientos de los estudiantes de postgrado en cursos dictados a través de la plataforma Moodle

Tarea de minería de datos: Asociación

Recolección de los datos:

- Para cada usuario matriculado en un curso dado, se logró recopilar su registro de actividades en la plataforma Moodle.

Preparación de los datos:

- Se construyeron variables en función de los porcentajes de utilización en Foros, Tareas, Recursos, Chats, SCORMS, Wiki y Cuestionario.
- Se encontraron ausencias en algunas variables; sin embargo, tenían una interpretación asociada con la interacción del estudiante en la plataforma (no utilizó el recurso).
- Selección de variables: Algoritmo InfoGainAttributeEval con el método Ranker. Se seleccionaron las variables:

%Foros

%Tareas

%Recursos

%Chats

Ejemplos de aplicaciones

- Se efectuó una discretización de las variables a intervalos equidistantes.

Intervalo	Etiqueta
[0, 0.2]	Muy bajo
(0.2, 0.4]	Bajo
(0.4, 0.6]	Medio
(0.6, 0.8]	Alto
(0.8, 1]	Muy alto



A cada intervalo se le asignó una etiqueta asociada al nivel de uso de la herramienta.

Minería de datos:

- **Tarea de minería de datos: Asociación**
- **Algoritmo: Apriori**
- **Medida de rendimiento: Soporte y confianza**
- **Confianza = 75%, Soporte = 50%**

Reglas encontradas:

R1: El 100% de los individuos en cuyos porcentajes tanto de entrega de tareas y de participación en chats fueron muy bajos también lo fue el porcentaje de participación en los foros.

R2: El 100% de los individuos con un porcentaje de tareas y de recursos muy bajos también tienen un porcentaje muy bajo de participación en los foros.

R3: El 100% de los estudiantes con un nivel bajo de entrega de tareas también tienen un porcentaje muy bajo de participación en los foros.

R4: Un 88% de los estudiantes cuya participación en los chats es muy baja también tienen un porcentaje muy bajo de participación en los foros.

R5: El 86% de los estudiantes cuyos niveles de participación es muy bajo tanto en foros como en chats también tienen un nivel muy bajo de entrega de tareas propuestas.

R6: El 75% de los individuos con un porcentaje muy bajo de participación en los chats también tienen un nivel muy bajo en la participación en los foros y en la entrega de tareas.

R7: El 75% de los individuos con un nivel muy bajo de participación en los chats también tienen un nivel muy bajo de entrega de tareas.

Ejercicio:

Aplique el proceso de minería de datos para realizar un análisis de asociación sobre el conjunto de datos “labor” (repositorio UCI). Este conjunto contiene datos de convenios colectivos alcanzados en el sector empresarial y de servicios personales, con un mínimo de 500 miembros. El objetivo es caracterizar estos acuerdos colectivos.

Observaciones:

- **Para tratar los valores ausentes los expertos consideran que una imputación por el valor medio resultaría adecuada. Sin embargo, sugieren hacerlo en aquellas variables que presentan hasta un 50% de ausencias. Con mas de 50% de ausencias los expertos consideran no incluir estas variables en el análisis**
- **Utilice un umbral para el soporte de 50% y una confianza de 80%.**

Tomando en cuenta el contexto del problema, realice una interpretación de las reglas generadas.