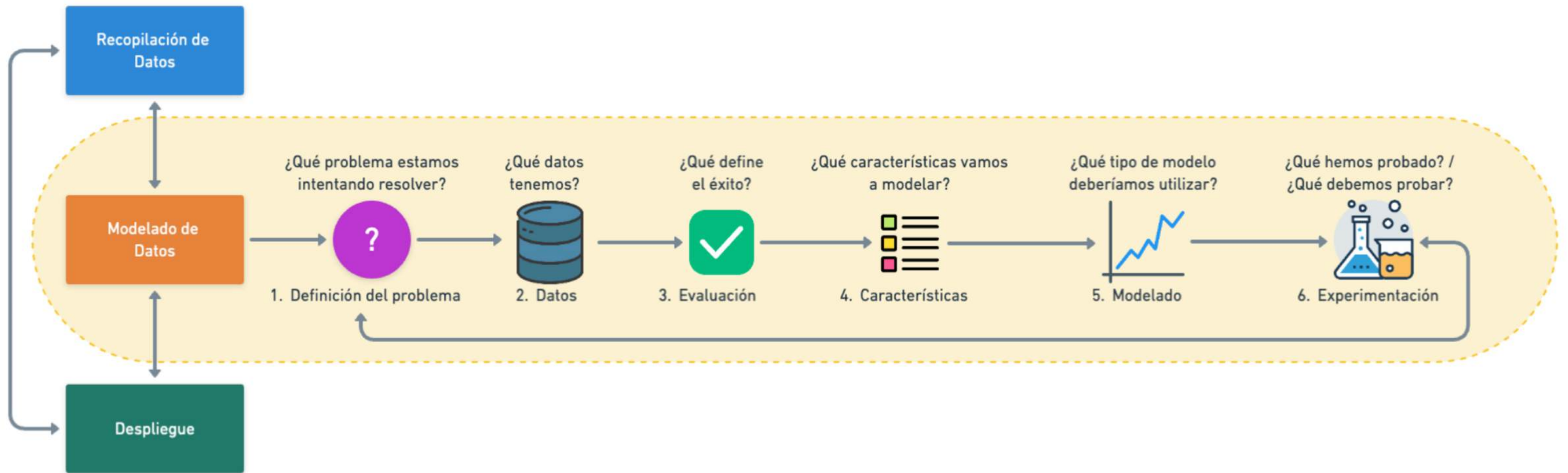


MACHINE LEARNING

MTRO. ALFONSO GREGORIO RIVERO DUARTE

FRAMEWORK



REGLAS DE ASOCIACIÓN

Los clientes que lo compraron, también compraron...

De esto va el tema de Aprendizaje con Reglas de Asociación!

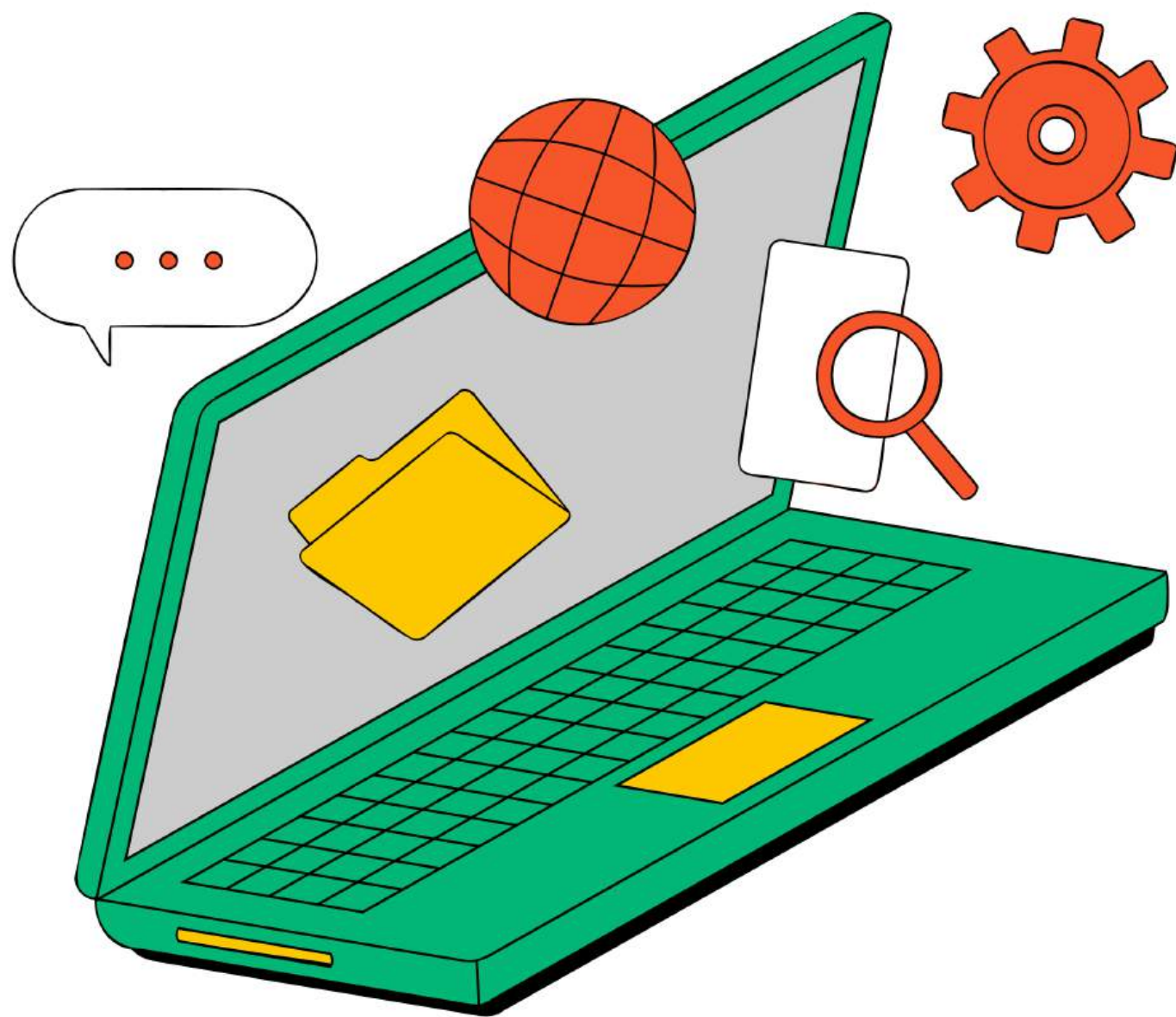


TÉCNICAS O ALGORITMOS

MODELOS DE CLUSTERING

1. Apriori

2. Eclat



REGLAS DE ASOCIACIÓN

MODELOS DE CLUSTERING

1. Idea de Apriori
2. Descripción del problema
3. Algoritmo de Apriori



IDEA DE APRIORI

APRIORI



@HarshaManoj

Alguna vez te ha pasado que vas a comprar algo y acabas comprando mucho más de lo que tenías previsto. Este es un fenómeno conocido como compra impulsiva y los grandes comercios se aprovechan de Machine Learning y más específicamente del algoritmo Apriori para asegurarse que los clientes compremos más.

Entender el algoritmo Apriori es fundamental para comprender muchas técnicas de análisis de la cesta de compra. Se utiliza para encontrar grupos de artículos que aparecen juntos con frecuencia en un conjunto de datos de compras. Este suele ser el primer paso para encontrar nuevas formas de promocionar la mercancía.

APRIORI

¿Cuál creéis que es la relación que existe entre todos, entre estos dos productos?



APRIORI

Bueno, cuenta la historia que en una tienda vendían una serie de productos, entre ellos estos dos y se hizo algún análisis acerca de cómo se comportaban los compradores con respecto a una serie de productos.

Mirando los datos, comprobaron que a través de miles y miles y miles de usuarios que pasaban por allí de transacciones que se llevaban a cabo, se dieron cuenta de que la gente que iba a comprar en ciertos momentos del día y en particular la gente que iba por la tarde entre las 6 y las 9 de la tarde, el supermercado en cuestión tendía a comprar estos dos productos juntos.

La gente que compraba pañales compraba cerveza o la gente que compraba cervezas compraba pañales. Lo cual era un poco extraño porque estos dos productos no están para nada relacionados.

APRIORI

Por qué alguien compraría cerveza cuando compra pañales o viceversa? En principio no existe una relación directa entre esos dos productos.

Por lo tanto algunas tiendas pueden decidir aprovecharse de este hecho, colocando las cervezas al lado de los pañales. Así el efecto contrario sería el padre va por pañales y termina llevando unas cervezas. Por tanto, ya está comprando otro item, lo cual es importante.

¿Pero cómo se llega hasta ahí? Pues ahí aparece el papel del algoritmo de a priori del que vamos a hablar a continuación. La gente que compró este producto también compró este otro.

Un caso famoso: el sistema de recomendación de películas de Netflix o Prime Video.

APRIORI

User ID	Películas que le han gustado
46578	Película1, Película2, Película3, Película4
98989	Película1, Película2
71527	Película1, Película2, Película4
78981	Película1, Película2
89192	Película2, Película4
61557	Película1, Película3

Reglas Significativas:

Película1  Película2

Película2  Película4

Película1  Película3

APRIORI

Transaction ID	Productos comprados
46578	Hamburguesas, Patatas, Verduras
98989	Hamburguesas, Patatas, Ketchup
71527	Verduras, Fruta
78981	Pasta, Fruta, Mantequilla, Verduras
89192	Hamburguesas, Pasta, Patatas
61557	Fruta, Zumo de Naranja, Verduras
87923	Hamburguesas, Patatas, Ketchup, Mayo

Reglas Significativas:	Hamburguesas	⇒	Patatas
	Verduras	⇒	Fruta
	Hamburguesas, Patatas	⇒	Ketchup

APRIORI

Y cómo funciona el algoritmo de a priori?

El algoritmo de apriori lo vamos a dividir en tres partes.

Soporte de la regla de asociación:

Se define como la cantidad de usuarios que ven la película o que transacciones que contienen un ítem. Usuarios que en general van a favor de lo que estáis evaluando. Dividido por la cantidad total de usuarios, sería como una especie de probabilidad.

¿Cuál es el soporte de una película M?

Pues son la cantidad de usuarios que ven la película dividido en la cantidad total de usuarios que tenemos en el dataset.

¿Cuál sería el soporte de un ítem de la cesta de la compra?

Sería el número de transacciones que contiene en ese ítem, dividido por el número total de transacciones que hay en el dataset

APRIORI

Recomendación de
Películas:

$$sop(\mathbf{M}) = \frac{|\text{usuarios que vieron } \mathbf{M}|}{|\text{usuarios}|}$$

Optimización de la
Cesta de la Compra:

$$sop(\mathbf{I}) = \frac{|\text{transacciones que contienen } \mathbf{I}|}{|\text{transacciones}|}$$

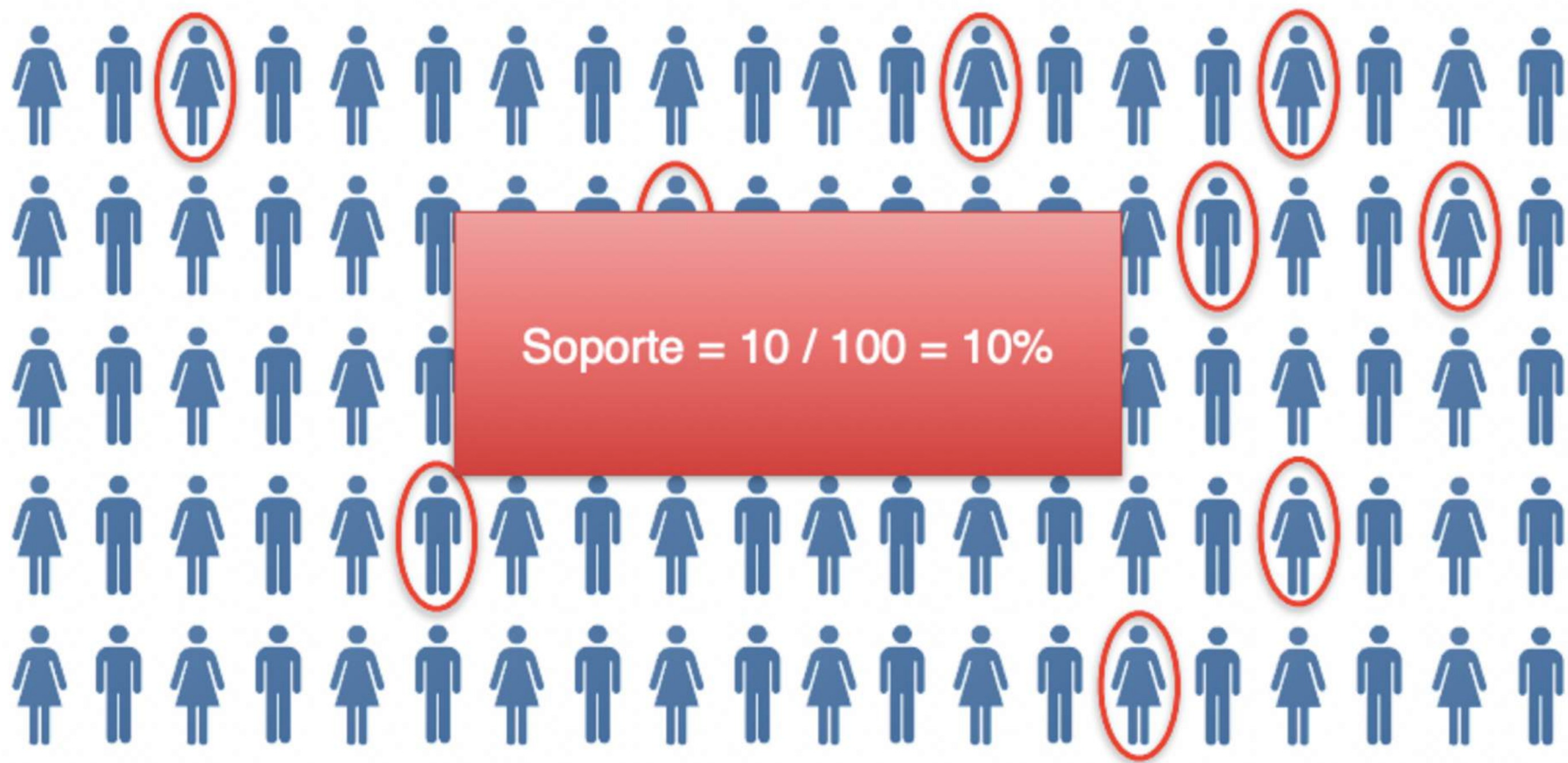
APRIORI

Imaginad que tenemos 100 personas organizadas en 5 filas y 20 columnas. Da igual si son hombres o mujeres, simplemente son seres humanos que queremos estudiar su comportamiento.

¿Cuántos de ellos están hablando de una película o han visto una película o le han dado me gusta a una película?

Eso significa que el soporte de ver la película es 10 sobre el total de 100 personas, o lo que es lo mismo, el 10 por ciento. Eso sería el soporte, sería el porcentaje o la proporción de usuarios a favor de un determinado evento.

APRIORI



APRIORI

Confianza:

¿Qué tan seguros estamos de que ocurren dos cosas a la vez?

Sería como una especie de probabilidad condicionada. En este caso la confianza de ver una película y entonces ver otra, pues se podría cuantificar como el número de usuarios que ven una película y ven la otra dividido el número total de usuarios que ven la primera de ellas.

Sería, por ejemplo, como una especie de hipótesis previa.

Por tanto, es una especie de cuantificación de cuántos usuarios que sé que hacen algo, también hacen otra cosa.

APRIORI

Recomendación de Películas:

$$\text{conf}(\mathbf{M}_1 \Rightarrow \mathbf{M}_2) = \frac{|\text{usuarios que vieron } \mathbf{M1} \text{ y } \mathbf{M2}|}{|\text{usuarios que vieron } \mathbf{M1}|}$$

Optimización de la Cesta de la Compra:

$$\text{conf}(\mathbf{I}_1 \Rightarrow \mathbf{I}_2) = \frac{|\text{transacciones que contienen } \mathbf{I1} \text{ y } \mathbf{I2}|}{|\text{transacciones que contienen } \mathbf{I1}|}$$

APRIORI

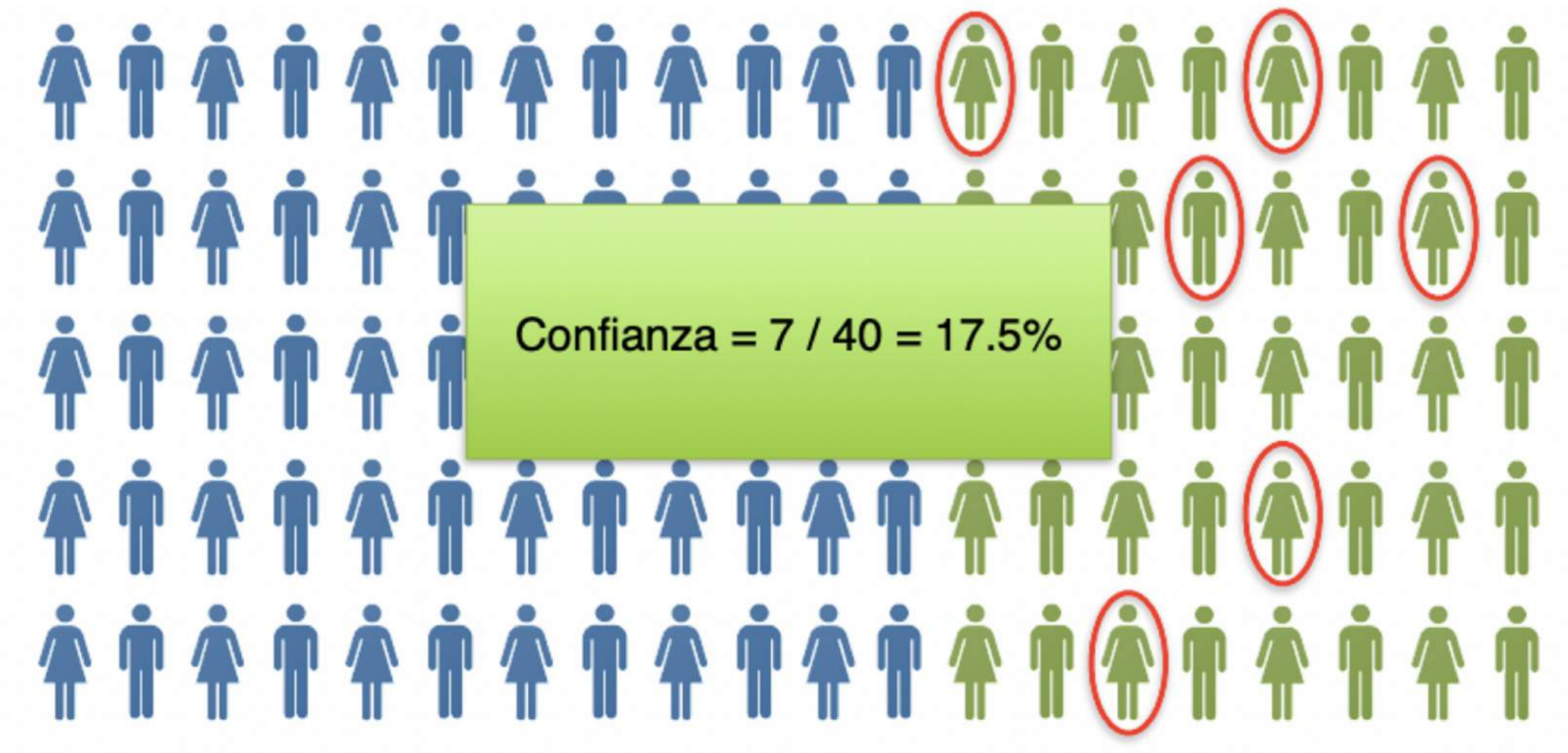
Los usuarios que están interesados o que sé que han visto la primera película, son los pintados de verde. 40 individuos.

Ahora, hay que encontrar de esas 40 personas, las que hayan visto la otra película X.

Por tanto, tenemos que 7 personas han visto las dos películas. Esa es la intersección de los dos sucesos y habría que dividirlo entre las 40 totales que han visto la primera.

Tomando eso en cuenta, tenemos una confianza de 17.5%

APRIORI



APRIORI

Lift:

El concepto de lift es bastante nuevo y es interesante. Se define como la confianza dividido por el soporte. Entonces calculamos la confianza. Entonces el lift es una forma de mejorar una respuesta aleatoria.

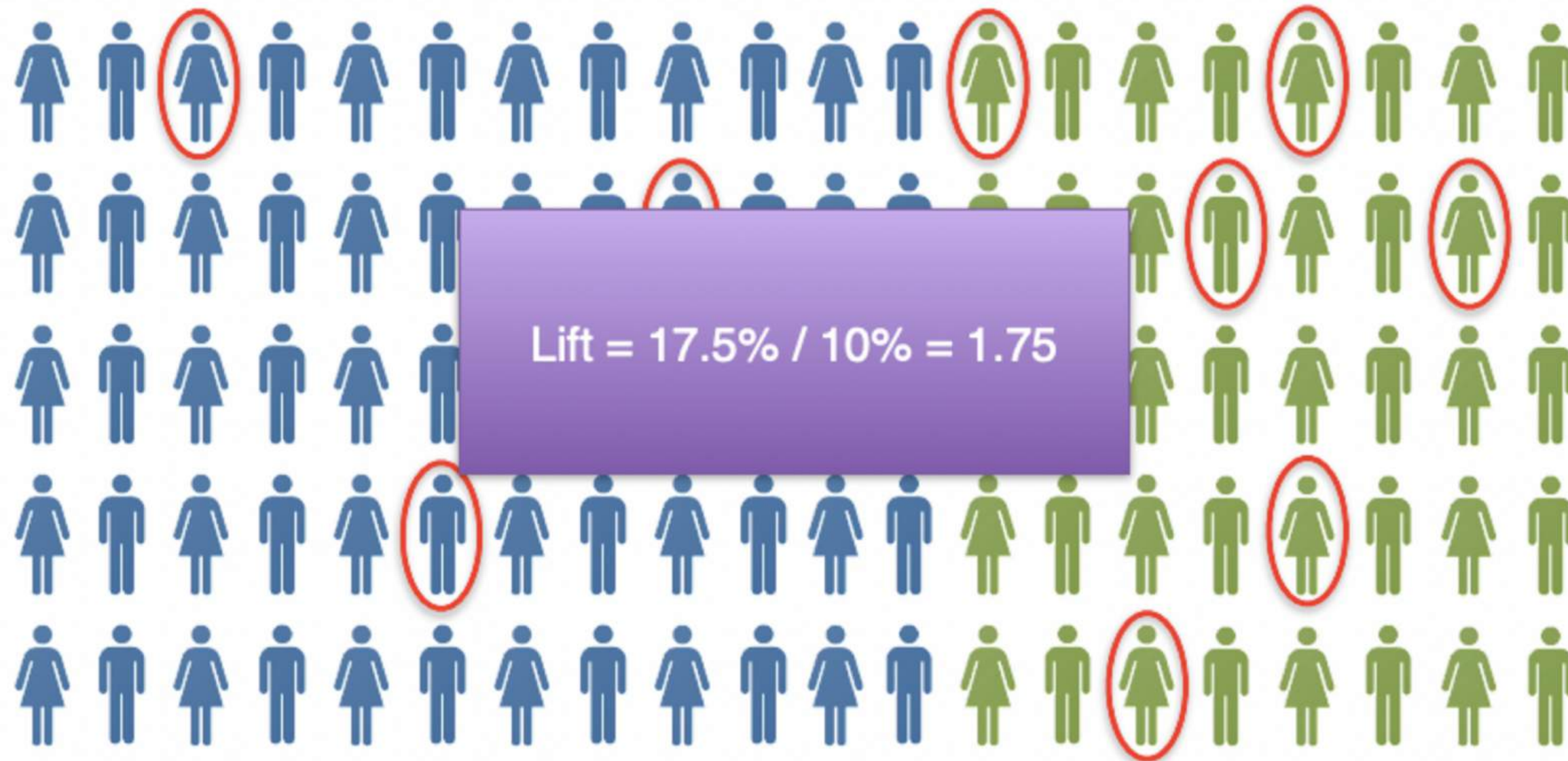
Recomendación
de Películas:

$$lift(\mathbf{M}_1 \Rightarrow \mathbf{M}_2) = \frac{conf(\mathbf{M}_1 \Rightarrow \mathbf{M}_2)}{sup(\mathbf{M}_2)}$$

Optimización de la
Cesta de la Compra:

$$lift(\mathbf{I}_1 \Rightarrow \mathbf{I}_2) = \frac{conf(\mathbf{I}_1 \Rightarrow \mathbf{I}_2)}{sup(\mathbf{I}_2)}$$

APRIORI



APRIORI

Establecemos un soporte mínimo y un nivel de confianza mínimo. Decidimos que es lo mínimo para nosotros para poder llevar a cabo el estudio de una recomendación. Este mínimo lo establecemos solo para evitar tener muchas, muchas, muchas recomendaciones en exceso. Solo nos interesa ejemplos que realmente tengan un poquito de significación o que el soporte no sea nulo.

Si no se establece unos límites al soporte y la confianza, el algoritmo se vuelve bastante lento.

Teniendo eso en mente, pues habrá que elegir todos los subconjuntos de transacciones con soporte superior al mínimo elegido.

APRIORI

Ahí está lo malo del algoritmo, que a lo mejor subconjuntos hay muchos. Podría ver que existieran muchos de todos los que hayan quedado, de todos los que superen ese soporte mínimo.

Lo que voy a hacer es elegir todas las reglas de asociación que se extraen de esos subconjuntos con un nivel de confianza superior al mínimo.

Finalizamos con un lift descendente.

APRIORI

Paso 1: Decidir un soporte y nivel de confianza mínimo



Paso 2: Elegir todos los subconjuntos de transacciones con soporte superior que el mínimo elegido



Paso 3: Elegir todas las reglas de estos subconjuntos con nivel de confianza superior al mínimo elegido



Paso 4: Ordenar todas las reglas anteriores por lift descendiente



DESCRIPCIÓN DEL PROBLEMA

APRIORI

Cómo optimizar las ventas de un centro comercial y entender cómo funciona el algoritmo de apriori.

Para hacer un trabajo de lujo, entendiendo el comportamiento de los clientes de ese centro comercial, la optimización de ventas en las tiendas ha aumentado considerablemente, agregando valor a la información que se tiene de los clientes.

Las empresas han estado utilizando las reglas de asociación para saber dónde colocar los productos en la tienda que colocar al lado de cada uno de los ítems.

Por poner un ejemplo muy simple si alguien compra unos cereales, es muy probable que esta persona también quiera comprar la leche para no tomarlo solos.



ALGORITMO

RETO

Reto:

Comencemos con:

- Active su ambiente local creado con miniconda en su equipo o su código de COLAB
- Abra el archivo [O9_Algoritmo_Apriori.ipynb](#)
- Ejecute el código y analícelo junto al instructor

Resultado Esperado:

	Left Hand Side	Right Hand Side	Support	Confidence	Lift
3	fromage blanc	honey	0.003333	0.245098	5.164271
0	light cream	chicken	0.004533	0.290598	4.843951
2	pasta	escalope	0.005866	0.372881	4.700812
8	pasta	shrimp	0.005066	0.322034	4.506672
7	whole wheat pasta	olive oil	0.007999	0.271493	4.122410
5	tomato sauce	ground beef	0.005333	0.377358	3.840659
1	mushroom cream sauce	escalope	0.005733	0.300699	3.790833
4	herb & pepper	ground beef	0.015998	0.323450	3.291994
6	light cream	olive oil	0.003200	0.205128	3.114710

PREGUNTAS Y RESPUESTAS

Mtro. Alfonso Gregorio Rivero Duarte

Senior Data Manager – CBRE

(+52) 5528997069

devil861109@gmail.com

<https://www.linkedin.com/in/alfonso-gregorio-rivero-duarte-139a9225/>

