# Reglas de Asociación entre Películas:

## Notebook Resumida

Hemos decidido trabajar con una regla de asociación de dos ítems, del modo A entonces B, es decir que nuestro objetivo es dada una película A vista por un usuario que otra única película proponerle para ver considerando que cierta cantidad de usuarios ya ha visto esa combinación (como son calificaciones se considera que el hecho de que haya visto ambas películas significa que ambas películas le gustaron).

Al trabajar con el dataset de Calificaciones, es decir las películas que ha visto y evaluado como buenas cada usuario, hemos tenido que bajar el support aceptable considerablemente, ya que al parecer es muy poco probable que en la lista se encuentre cierta regla de A entonces B en particular, posiblemente por la cantidad de calificaciones existentes. En el caso práctico, aun bajando a 0,001 el mínimo aceptable hemos tenido que disminuir las transacciones a apenas 1000 porque con más filas el algoritmo no devolvía resultados.

No sucede lo mismo con confidence que, si bien usamos un valor mínimo de prueba vemos que los valores obtenidos son 0,5 y 1 en la mayoría de los casos. Es decir, que es muy probable que un usuario vea una película B dado que vio A. Vale aclarar que estos valores altos se deben en este caso a las pocas reglas que devolvió el algoritmo, si vemos el primer ejemplo de conf =1 en la notebook se aprecia que la película analizada figura solo una vez en toda la muestra, por ello es lógico que el resultado sea del 100%.

Por ejemplo:

* {Matrix Reloaded, The} -> {Batman Forever} (conf: 1.000, supp: 0.001, lift: 986.000, conv: 998985801.217)

Conf = 1, significa que todo usuario que vio Matrix Reloaded, The entonces vio Batman Forever.

(Ver ejemplo 1 en Notebook Resumen)

* {Galaxy Quest} -> {Steal This Film} (conf: 0.333, supp: 0.001, lift: 328.667, conv: 1.498)

Conf = 0,33, significa que el 33% de los usuario que vieron Galaxy Quest entonces vieron Steal This Film.

(Ver ejemplo 2 en Notebook Resumen)

En estos casos lift es muy elevado por esto mismo de tener una muestra de reglas muy reducida, lo que nos dice está métrica es que la probabilidad de ver las películas A y B es tantas veces más que ver solo A, por ende si es mayor a 1, como es en estos casos, significa que es altamente probable que A y B se vean juntas entonces.

Nota:

No solo se trabajó con pocas calificaciones para demostrar que incluso con support mínimo no se obtienen casi resultados sino por problemas de performance, resultó difícil con mis recursos manipular estos datos.

Este mismo análisis podría hacerse por género además de por película.

## Notebook Ampliada

No fue posible ejecutar con más calificaciones por los recursos disponibles.