

Análisis de Reglas de Asociación en Compras Online



Gerard Alba Soler

Introducción

El objetivo de este trabajo es implementar el algoritmo no-supervisado Apriori para encontrar relaciones estructuradas entre elementos del dataset de compras online. Gracias a este algoritmo identificamos patrones de compra lo cual puede ser muy útil para personalizar más la experiencia de los clientes, con recomendaciones más específicas, diseñar estrategias de marketing y optimizar las opciones de envío.

Datos

Los datos se han obtenido del dataset público de compras online proporcionado por el profesor, en este dataset nos encontramos con transacciones de diferentes países. Cada entrada en el dataset representa un producto comprado en una transacción, esto nos permite ver información como la descripción del producto, la cantidad, el precio, el número de factura, el país del cliente, entre otros.

Df.Head()

	InvoiceNo	StockCode	Description	Quantity	\
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	
1	536365	71053	WHITE METAL LANTERN	6	
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	

	InvoiceDate	UnitPrice	CustomerID	Country
0	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Metodología

1. **Preprocesamiento de Datos:** Se eliminaron las filas con valores nulos y se filtraron las transacciones por país.
2. **Codificación de Datos:** Los datos se convirtieron en una matriz binaria, donde cada fila representa una transacción y cada columna un producto. Si un producto está en una transacción, se codifica como 1, de lo contrario, como 0.
3. **Implementación del Algoritmo Apriori:** Se utilizó la librería **mlxtend** para implementar el algoritmo Apriori y encontrar conjuntos de ítems frecuentes y reglas de asociación para cada país.
4. **Análisis de Reglas:** Se analizaron las reglas generadas para identificar patrones y hacer recomendaciones.

```
# Eliminar filas con valores nulos
df.dropna(inplace=True)
# Filtrar las transacciones por país (por ejemplo, Francia)
df_france = df[df['Country'] == 'France']
# Convertir los datos en una matriz binaria
basket = (df_france.groupby(['InvoiceNo', 'Description'])['Quantity']
          .sum().unstack().reset_index().fillna(0)
          .set_index('InvoiceNo'))

def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1
    (parameter) use_colnames: bool

basket_sets = basket.applymap(encode_units)
# Encontrar conjuntos de ítems frecuentes
frequent_itemsets = apriori(basket_sets, min_support=0.07, use_colnames=True)
# Encontrar reglas de asociación
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
# Ordenar las reglas por confianza
rules_sorted = rules.sort_values(['confidence', 'lift'], ascending=[False, False])
# Mostrar las primeras 10 reglas
print(rules_sorted.head(10))
# Estadísticas descriptivas
print(rules.describe())
# Visualizar las reglas usando un gráfico de dispersión
plt.scatter(rules['support'], rules['confidence'], alpha=0.5)
plt.xlabel('support')
plt.ylabel('confidence')
plt.title('Support vs Confidence')
plt.show()
```

Resultados

Francia:

Principales hallazgos:

1. Productos Relacionados con Fiestas y Eventos:

- Una combinación interesante es la de "SET/20 RED RETROSPOT PAPER NAPKINS" y "SET/6 RED SPOTTY PAPER PLATES". Se logra observar que, cuando estos productos se compran juntos, por lo general el cliente además compra de "SET/6 RED SPOTTY PAPER CUPS". Esta combinación sugiere que los clientes que compran servilletas y platos con un diseño de lunares rojos también tienden a comprar vasos con un diseño similar. Esto podría asociarse con preparativos para fiestas o eventos.

2. Productos de Envío:

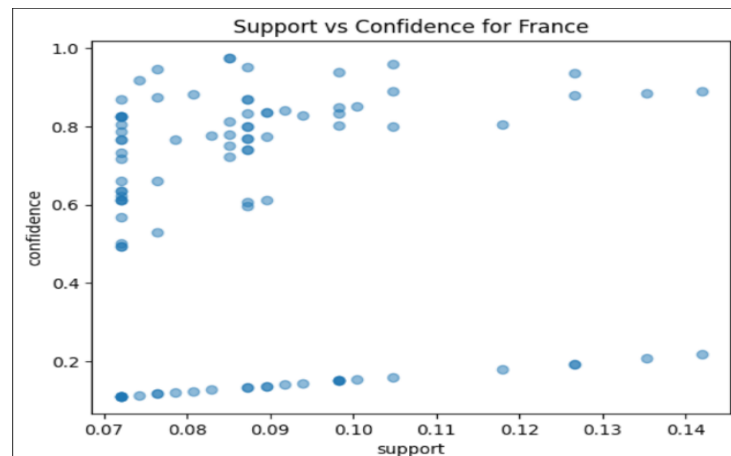
- Varios productos, como "TEA PARTY BIRTHDAY CARD", "STRAWBERRY LUNCH BOX WITH CUTLERY", "ROUND SNACK BOXES SET OF 4 WOODLAND", "ALARM CLOCK BAKELIKE RED", y "RABBIT NIGHT LIGHT", muestran una fuerte asociación con "POSTAGE". Con esto podemos inferir que estos productos tienen tendencia a ser enviados, en lugar de ser comprados en tienda física.

3. Productos Decorativos:

- La asociación entre "SET/6 RED SPOTTY PAPER PLATES" y "SET/6 RED SPOTTY PAPER CUPS" podemos inferir que estos productos decorativos son populares juntos, Podría ser debido a su diseño similar o por el tipo de evento en el que se utilizan.

Estadísticas:

- La métrica de "lift" tiene un rango [1.27 - 8.93], lo que expresa que algunas de estas reglas son hasta 8.93 veces más probables de lo esperado, con estos resultados inferimos que hay una fuerte relación entre los productos antecedente y consecuente.
- La métrica de "confidence" tiene un valor máximo de 0.91, este valor señala que existen reglas con una alta confianza, para ser más preciso en el 91% de las transacciones que contienen el antecedente, también se encuentra el consecuente.
- Con la métrica de "support" encontramos que la proporción de transacciones en el dataset que contienen el ítemset. En este caso, tiene un rango [0.07 - 0.16], gracias a esto conocemos que estas reglas se aplican a un porcentaje significativo de transacciones.



Alemania:

Principales hallazgos:

1. Productos Relacionados con Alimentos y Snacks:

- Una combinación destacada es la de "ROUND SNACK BOXES SET OF 4 FRUITS" y "ROUND SNACK BOXES SET OF 4 WOODLAND". Observamos que la compra en conjunto de estos productos se puede asociar también con "POSTAGE". Esto sugiere que estos contenedores de snacks son populares entre los clientes alemanes y a menudo se envían juntos.

2. Productos de Cuidado Personal y Salud:

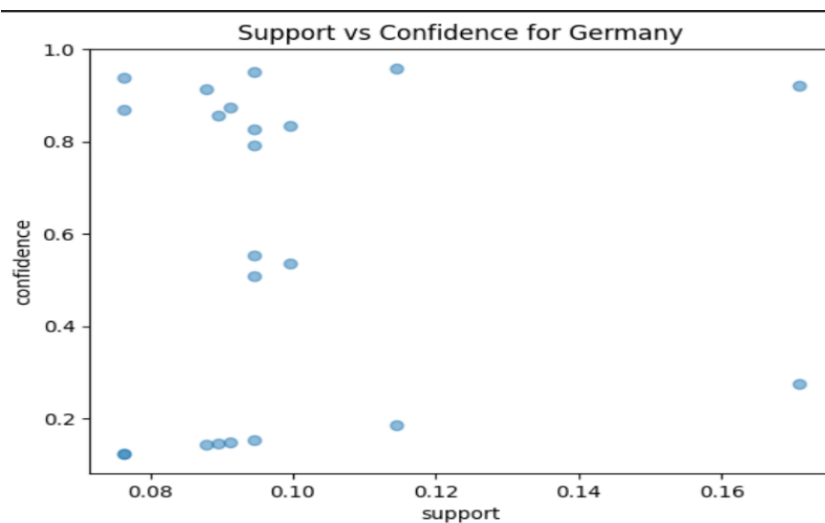
- "PLASTERS IN TIN SPACEBOY", "PLASTERS IN TIN CIRCUS PARADE" y "PLASTERS IN TIN WOODLAND ANIMALS" también muestran una fuerte asociación con "POSTAGE". Como resultado de este análisis observamos que estos productos de cuidado personal son a menudo enviados, lo que puede sugerir que son populares como regalos o que no están fácilmente disponibles en tiendas físicas locales o incluso que a los clientes les parece más cómodo comprarlos de esta manera.

3. Productos Decorativos:

- "WOODLAND CHARLOTTE BAG" y "REGENCY CAKESTAND 3 TIER" también muestran una fuerte asociación con "POSTAGE". Al igual que los otros productos existen razones por las cuales los clientes optan por esta opción de compra y sería bueno darle un mayor análisis.

Estadísticas:

- La métrica de "lift" tiene un rango [1.27 - 4.63], Esto muestra que algunas de estas reglas son hasta 4.63 veces más probables de lo esperado, Este tipo de resultados sugieren una fuerte relación entre los productos antecedente y consecuente.
- La métrica de "confidence" tiene un valor máximo de 0.77, Esto nos apunta a que hay reglas con una alta confianza, en otras palabras, en el 77% de las transacciones que contienen el antecedente, también se encuentra el consecuente.
- Cuando hablamos de la métrica de 'support', nos referimos a cuántas de nuestras transacciones totales incluyen un conjunto específico de ítems. En nuestro caso, este rango varía entre [0.08 - 0.18]. Esto significa que estas reglas no son solo aplicables a unas pocas transacciones, sino que son relevantes para una parte considerable de todas las transacciones que tenemos en nuestro dataset.



España:

Principales hallazgos:

1. Productos de Fiesta y Decoración:

- Una combinación destacada es la de "PACK OF 72 RETROSPOT CAKE CASES", "6 RIBBONS RUSTIC CHARM", y "ASSORTED COLOUR BIRD ORNAMENT". Estos productos, cuando se compran juntos los podemos observar como una venta "POSTAGE". Esto señalaría que los clientes que efectúan la compra productos decorativos y de fiesta menudo se envían juntos.

2. Productos de Jardín y Decoración:

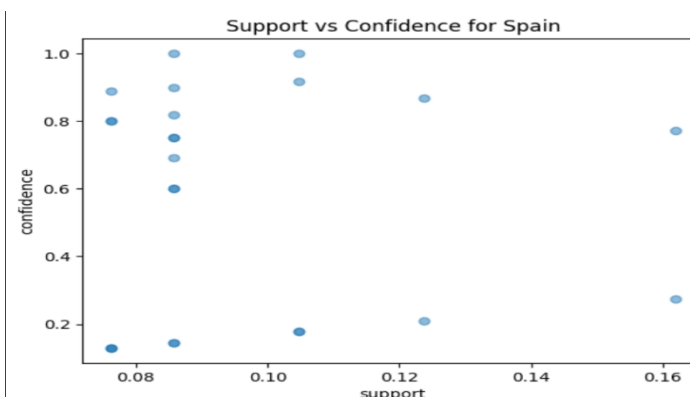
- "CLASSIC METAL BIRDCAGE PLANT HOLDER" muestra una fuerte asociación con "POSTAGE". Esto indica que este producto decorativo para jardín se vende más de forma remota para envíos.

3. Productos de Cocina y Mesa:

- "SET/20 RED RETROSPOT PAPER NAPKINS" y "SET OF 72 RETROSPOT PAPER DOILIES" también tienen una fuerte relación con "POSTAGE". Debido a esto es bueno considerar estrategias de envío gratis para promocionar los productos

Estadísticas:

- La métrica de "lift" tiene un rango que varía entre [1.27 - 6.05], esto se traduce en que algunas de estas reglas son hasta 6.05 veces más probables de lo que se esperaría inicialmente, lo cual nos lleva a pensar que existe una relación bastante fuerte entre los productos que aparecen antes y después en nuestras transacciones.
- La métrica de "confidence" tiene un valor máximo de 0.71, Esto nos señala que existen reglas con una confianza bastante alta. Para ponerlo en perspectiva, en el 71% de las ocasiones que los productos del antecedente son adquiridos, los del consecuente también se añaden a la compra. Es como si estos productos estuvieran destinados a ser comprados juntos.
- Cuando hablamos de la métrica de 'support', estamos refiriéndonos a cuántas de las transacciones totales en nuestro conjunto de datos incluyen un conjunto específico de ítems. En este caso, encontramos que este rango se sitúa entre 0.08 y 0.20. Esto nos dice que estas reglas no son solo aplicables a unas pocas transacciones, sino que son relevantes para una parte considerable de todas las transacciones que tenemos en nuestro conjunto de datos.



Conclusión Consolidada de Reglas de Asociación para Francia, Alemania y España

Tras analizar las reglas de asociación de las compras en Francia, Alemania y España, nos encontramos con patrones y tendencias distintivas que reflejan las preferencias y comportamientos de compra en cada país y sus respectivas preferencias:

1. **Francia:** Los clientes franceses muestran una inclinación hacia productos de fiesta y decoración en este dataset, como los platos y vasos de papel con diseño "Red Spotty". Además, hay una fuerte tendencia a comprar productos relacionados con la decoración y el envío, estos artículos podrían ser regalos o compras para ocasiones especiales.
2. **Alemania:** Los clientes alemanes parecen valorar los productos relacionados con la comida y la decoración del hogar. Las cajas de snack "Round" y los productos con diseño "Woodland" son especialmente populares en el dataset. Además, hay una fuerte asociación con productos de primeros auxilios, como las tiritas con diseños temáticos. Los clientes alemanes buscan tanto funcionalidad como estética en sus compras.
3. **España:** Los patrones de compra en España en este dataset reflejan una mezcla de decoración y productos de fiesta. Los productos "Retrosport", "Ribbons Rustic Charm" y "Bird Ornament" son especialmente populares. Al igual que en Francia, hay una tendencia a comprar estos productos junto con el envío esto genera oportunidades de negocio o promociones relacionadas.

Implicaciones Estratégicas:

- **Promociones Personalizadas:** Las empresas pueden diseñar promociones y descuentos específicos para cada país basándose en estos patrones de compra. Por ejemplo, ofrecer descuentos en productos de fiesta en Francia o promociones en productos de decoración del hogar en Alemania.
- **Recomendaciones de Productos:** Las plataformas de comercio electrónico pueden utilizar estas reglas para hacer recomendaciones de productos más precisas a los clientes de cada país, mejorando así la experiencia del cliente y aumentando las ventas.
- **Gestión de Inventario:** Estos insights pueden ayudar a las empresas a gestionar mejor su inventario, asegurando que los productos populares estén siempre disponibles en cada país.

Conclusión Final:

El análisis de reglas de asociación ha proporcionado insights valiosos sobre las preferencias de compra en Francia, Alemania y España. Estos patrones reflejan las culturas y comportamientos únicos de cada país. Al actuar sobre estos insights, las empresas pueden ofrecer una experiencia de compra más personalizada y eficiente, lo que a su vez puede conducir a una mayor lealtad del cliente y un aumento en las ventas.