

- **Ejercicio Reglas de Asociación Market Basket**

Para el análisis en el Reino Unido, utilizamos un soporte mínimo del 3%, lo que significa que solo se considerarán los productos o combinaciones que aparezcan en al menos el 3% de todas las transacciones. Este umbral se eligió para asegurar que solo se incluyan interacciones con una frecuencia significativa. Sin embargo, en el conjunto de datos del Reino Unido, no hay suficientes combinaciones que cumplan con este criterio, lo que podría limitar la cantidad de reglas de asociación generadas.

- Para el ítem (6 RIBBONS RUSTIC CHARM), el soporte es 0.045800, lo que significa que este producto aparece en aproximadamente el 4.58% de todas las transacciones.
- Para el conjunto de ítems (JUMBO BAG RED RETROSPOT, JUMBO BAG PINK POLKA...), el soporte es 0.042051, lo que significa que esta combinación de productos aparece en aproximadamente el 4.21% de todas las transacciones.

	support	itemsets
0	0.045800	(6 RIBBONS RUSTIC CHARM)
1	0.031123	(60 CAKE CASES VINTAGE CHRISTMAS)
2	0.040336	(60 TEATIME FAIRY CAKE CASES)
3	0.046925	(ALARM CLOCK BAKELIKE GREEN)
4	0.035140	(ALARM CLOCK BAKELIKE PINK)
...
126	0.030534	(JUMBO BAG RED RETROSPOT, JUMBO BAG BAROQUE B...
127	0.042051	(JUMBO BAG RED RETROSPOT, JUMBO BAG PINK POLKA...
128	0.035194	(JUMBO BAG RED RETROSPOT, JUMBO SHOPPER VINTAG...
129	0.037390	(JUMBO STORAGE BAG SUKI, JUMBO BAG RED RETROSPOT)
130	0.032516	(LUNCH BAG RED RETROSPOT, LUNCH BAG BLACK SKU...

Algunos productos como (ALARM CLOCK BAKELIKE GREEN) tienen un soporte más alto, lo que indica que son productos populares y se venden con bastante frecuencia de manera individual.

Con conjuntos como (JUMBO BAG RED RETROSPOT, JUMBO BAG BAROQUE), esto indica que hay una tendencia de estos productos a comprarse juntos. El soporte de estos conjuntos sugiere con qué frecuencia aparecen estas combinaciones en las transacciones.

Utilizando el algoritmo Apriori

(ALARM CLOCK BAKELIKE GREEN) tiene un soporte consecuente de 0.046925, apareciendo en el 4.69% de las transacciones.

El soporte de la regla completa es 0.030159, lo que significa que el conjunto (ALARM CLOCK BAKELIKE RED) y (ALARM CLOCK BAKELIKE GREEN) aparecen juntos en aproximadamente el 3.02% de las transacciones.

La confianza es 0.605376, lo que significa que cuando (ALARM CLOCK BAKELIKE RED) se compra, hay un 60.54% de probabilidad de que también se compre (ALARM CLOCK BAKELIKE GREEN). Con un “lift” de 12.900874 se indica que la compra conjunta de estos dos productos es 12.9 veces más probable de lo que sería si fueran independientes, el “leverage” es 0.027821, lo que indica que la coocurrencia de estos productos es ligeramente mayor de lo esperado, el “conviction” es 2.415149, lo que indica una relación relativamente fuerte entre los productos, la métrica de “Zhang” es 0.970852, lo que indica una alta calidad de la regla.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.049818	0.046925	0.030159	0.605376	12.900874	0.027821	2.415149	0.970852
1	(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	0.046925	0.049818	0.030159	0.642694	12.900874	0.027821	2.659296	0.967905
2	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.037658	0.050032	0.030909	0.820768	16.404818	0.029024	5.300218	0.975789
3	(GREEN REGENCY TEACUP AND SAUCER)	(PINK REGENCY TEACUP AND SAUCER)	0.050032	0.037658	0.030909	0.617773	16.404818	0.029024	2.517724	0.988499
4	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.050032	0.051264	0.037551	0.750535	14.640537	0.034986	3.803087	0.980766

Alfonso Camarillo Nuñez

5	(ROSES REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.051264	0.050032	0.037551	0.732497	14.640537	0.034986	3.551247	0.982040
6	(JUMBO BAG RED RETROSPOT)	(JUMBO BAG BAROQUE BLACK WHITE)	0.103814	0.048747	0.030534	0.294118	6.033613	0.025473	1.347609	0.930903
7	(JUMBO BAG BAROQUE BLACK WHITE)	(JUMBO BAG RED RETROSPOT)	0.048747	0.103814	0.030534	0.626374	6.033613	0.025473	2.398615	0.877013
8	(JUMBO BAG RED RETROSPOT)	(JUMBO BAG PINK POLKADOT)	0.103814	0.062085	0.042051	0.405057	6.524245	0.035605	1.576478	0.944810
9	(JUMBO BAG PINK POLKADOT)	(JUMBO BAG RED RETROSPOT)	0.062085	0.103814	0.042051	0.677308	6.524245	0.035605	2.777218	0.902774
10	(JUMBO BAG RED RETROSPOT)	(JUMBO SHOPPER VINTAGE RED PAISLEY)	0.103814	0.060692	0.035194	0.339009	5.585724	0.028893	1.421061	0.916073
11	(JUMBO SHOPPER VINTAGE RED PAISLEY)	(JUMBO BAG RED RETROSPOT)	0.060692	0.103814	0.035194	0.579876	5.585724	0.028893	2.133149	0.874018
12	(JUMBO STORAGE BAG SUKI)	(JUMBO BAG RED RETROSPOT)	0.060531	0.103814	0.037390	0.617699	5.950055	0.031106	2.344190	0.885537
13	(JUMBO BAG RED RETROSPOT)	(JUMBO STORAGE BAG SUKI)	0.103814	0.060531	0.037390	0.360165	5.950055	0.031106	1.468299	0.928305
14	(LUNCH BAG RED RETROSPOT)	(LUNCH BAG BLACK SKULL.)	0.074566	0.065138	0.032516	0.436063	6.694431	0.027658	1.657742	0.919160
15	(LUNCH BAG BLACK SKULL.)	(LUNCH BAG RED RETROSPOT)	0.065138	0.074566	0.032516	0.499178	6.694431	0.027658	1.847829	0.909891

Y con la reglas de asociación filtradas tenemos que el support es de 0.030909 esta métrica muestra que ambos productos aparecen juntos en el 3.09% de las transacciones, tenemos que la confianza de la regla es del 82.08%, lo que significa que, cuando alguien compra "PINK REGENCY TEACUP AND SAUCER", hay un 82.08% de probabilidad de que también compre "GREEN REGENCY TEACUP AND SAUCER". Esta regla es muy fuerte y sugiere que los clientes que compran el set de té "PINK REGENCY TEACUP AND SAUCER" tienen una alta probabilidad de también comprar el set "GREEN REGENCY TEACUP AND SAUCER".

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
2	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.037658	0.050032	0.030909	0.820768	16.404818	0.029024	5.300218	0.975789

- **Ejercicio Árbol de decisión**

Al analizar la data de la base de datos de Diabetes tenemos los porcentajes de positivos y negativos.

```
diabetes
neg    0.67
pos    0.33
```

Al dividir los datos para probar y entrenar nos queda el siguiente porcentaje

```
proporción target entrenamiento
diabetes
0          0.67
1          0.33
Name: proportion, dtype: float64
-----
proporción target test
diabetes
0          0.67
1          0.33
```

Como respuesta del modelo tenemos

```
===== Con Tuning =====
Indicador
GINI    0.577974
AUC     0.788987
KS       0.577974
Test:
Indicador
GINI    0.336904
AUC     0.668452
KS       0.336904
===== Sin Tuning =====
Train:
Indicador
GINI    1.0
AUC     1.0
KS       1.0
Test:
Indicador
GINI    0.259656
AUC     0.629828
KS       0.259656

===== Con Tuning =====
Accuracy - Train: 0.7956204379562044
Accuracy - Test : 0.6864406779661016
===== Sin Tuning =====
Accuracy - Train Sin Tuning: 1.0
Accuracy - Test Sin Tuning : 0.6694915254237288
```

Con tuning

GINI: 0.336904 indicando que el modelo tiene menos capacidad de discriminación en los datos de prueba.

AUC: 0.668452 está por encima de 0.5, lo que significa que el modelo tiene una capacidad de clasificación razonable.

Alfonso Camarillo Nuñez

KS: 0.336904: Al ser bajo, indicar una menor efectividad del modelo en los datos de prueba comparado con los datos de entrenamiento.

Con tuning el modelo es más robusto y generaliza mejor a nuevos datos. Aunque los valores de GINI, AUC, y KS son menores en el conjunto de prueba comparado con el de entrenamiento, siguen indicando un modelo razonablemente efectivo.

La glucosa emerge como la variable más importante y con mayor influencia en la detección de la diabetes, tal como se esperaba. Es la única característica que contribuye de manera significativa a las predicciones, mientras que las demás variables no muestran una influencia relevante en este contexto.

