# Exercices pour le thème 2: les règles d’association

## Exercice 1

Répondre aux questions suivantes à l’aide du tableau de support suivant :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Beurre** | **Vin** | **Fromage** | **Pain** |
| **Beurre** | 30.00% |  |  |  |
| **Vin** | 0.00% | 25.00% |  |  |
| **Fromage** | 5.00% | 15.00% | 65.00% |  |
| **Pain** | 30.00% | 15.00% | 50.00% | 80.00% |

1. Quel item est le plus populaire?
2. Construisez le tableau de la confiance
3. Quelle est la règle ayant la confiance la plus élevée? Interprétez sa valeur.
4. Construisez le tableau du lift
5. Quel est le meilleur lift obtenu? Interprétez sa valeur.
6. Quel est le pire lift obtenu? Interprétez sa valeur.
7. Quels sont les deux items qui semblent être les plus indépendants l’un de l’autre?

## Exercice 2

Reprendre la base de données du cas Boulangerie (fichier 30-650-boulangerie-DATA.xlsx) et répondre aux questions suivantes.

1. Pour la paire d’items (tarte au citron, gâteau au citron), construisez le tableau croisé dynamique et calculez ensuite les supports individuels, le support conjoint, les confiances et le lift. Interprétez ensuite les valeurs.

À l’aide du gabarit « Association », calculez les supports, confiances et lifts pour toutes les règles à deux items et répondre aux questions suivantes :

1. À l’aide des indicateurs numériques, analysez les deux règles associées à la paire d’items (croissant au fromage, croissant aux amandes). Que pouvez-vous conclure quant à l’association entre ces deux items?
2. À l’aide des indicateurs numériques, analysez les deux règles associées à la paire d’items (café, thé vert). Que pouvez-vous conclure quant à l’association entre ces deux items?
3. Si un client a déjà choisi un capuccino, quel item devriez-vous lui suggérer? Quelle est alors la probabilité qu’il achète cet item proposé? Quelle aurait été la probabilité qu’il achète ce second item sans savoir qu’il achetait au préalable un capuccino?

**Solutions**

## Exercice 1

1. On regarde l’item ayant le support individuel le plus élevé. C’est donc le pain dont la probabilité d’achat est estimée à 80% selon ces données.
2. La confiance s’obtient en prenant le ratio du support de la règle sur le support de l’antécédent, i.e.,

**Confiance (X => Y)** =

On obtient donc le tableau de confiance suivant :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Beurre** | **Vin** | **Fromage** | **Pain** |
| **Beurre** |  | 0.00% | 16.67% | 100.00% |
| **Vin** | 0.00% |  | 60.00% | 60.00% |
| **Fromage** | 7.69% | 23.08% |  | 76.92% |
| **Pain** | 37.50% | 18.75% | 62.50% |  |

1. La confiance maximale est atteinte pour la règle (Beurre = > Pain) qui est de 100%. Il s’agit d’une estimation de la probabilité d’acheter du pain sachant que du beurre est acheté. Ainsi, le client achètera nécessairement du pain s’il achète au préalable du beurre.
2. Le lift s’obtient en prenant le ratio de la confiance de la règle sur le support du conséquent.

lift (X => Y) =

On obtient donc le tableau de lift suivant :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Beurre** | **Vin** | **Fromage** | **Pain** |
| **Beurre** |  |  |  |  |
| **Vin** | 0.00 |  |  |  |
| **Fromage** | 0.26 | 0.92 |  |  |
| **Pain** | 1.25 | 0.75 | 0.96 |  |

1. Le meilleur lift s’obtient pour l’association entre le beurre et le pain, i.e., pour les deux règles : Beurre = > Pain et Pain => Beurre. La valeur de 1.25 indique pour la règle (Beurre => Pain) que la probabilité d’acheter du pain augmente d’un facteur de 1.25 lorsque le client achète du beurre. En effet, la probabilité d’acheter du pain passe de 80% à 100% suite à l’achat du beurre. Puisque le lift est symétrique, il en est de même pour la règle inverse (Pain => Beurre). En effet, la probabilité d’acheter du beurre augmente aussi d’un facteur de 1.25 passant de 30% à 37.5% suite à l’achat du pain.
2. Le pire lift s’obtient pour l’association entre le beurre et le vin, i.e., pour les deux règles : Beurre = > Vin et Vin => Beurre. La valeur de 0 indique que les clients n’achètent jamais les deux produits en même temps. Ainsi, si le client achète du beurre, il n’achètera pas de vin et vice versa. On le constate aussi par le fait que la confiance est nulle dans les deux cas.
3. Le pain et le fromage sont les deux items qui sont les plus près d’être indépendants, lift plus près de 1. En effet, pour les deux règles issues de cette paire d’items, on constate que la confiance est légèrement inférieure au support du conséquent (ratio de 0.96). En effet, la probabilité d’acheter du pain passe de 80% à 76.92% suite à l’achat du fromage et celle du fromage passe de 65% à 62.5% suite à l’achat du pain.

## Exercice 2

1. Le tableau croisé est le suivant :



Où lignes = gâteau au citron (GC) et colonnes = tarte au citron (TC)

Les deux règles sont : (TC => GC) et (GC => TC)

Support (GC) =

La probabilité qu’un client achète du gâteau au citron est estimée à 9,4% (pourcentage de transactions où le gâteau au citron est présent).

Support (TC) =

La probabilité qu’un client achète de la tarte au citron est estimée à 7,83% (pourcentage de transactions où la tarte au citron est présente).

Support (TC => GC) = Support (GC=>TC) =

La probabilité qu’un client achète les deux items est estimée à 3,71% (pourcentage de transactions où la tarte au citron est présente).

Confiance (TC => GC) =

La probabilité qu’un client achète du gâteau citron sachant qu’il a acheté de la tarte au citron est estimée à 47,46%.

Confiance (GC => TC) =

La probabilité qu’un client achète de la tarte au citron sachant qu’il a acheté du gâteau au citron est estimée à 39,53%.

Lift (TC => GC) = Lift (GC=>TC) =

On a une association positive. Le fait de savoir que le client a acheté un des deux items fait plus que quintupler la probabilité qu’il achète l’autre item.

1. Les deux règles associées aux items croissant au fromage (CF) et croissant aux amandes (CA) sont : (CF => CA) et (CA => CF). Tout d’abord, on constate que le lift est de 1.03, ce qui laisse croire que les deux items sont quasi-indépendants. En effet, on constate que la confiance de (CF => CA) ≈ 5,19% et que le support de CA ≈ 5,04%. Ainsi, la probabilité d’acheter des croissants aux amandes est très peu affectée par le fait d’avoir acheté des croissants au fromage. Il en est de même pour la règle inverse puisque confiance (CA => CF) ≈ 8,77% et support (CF) ≈ 8,51%.
2. Les deux règles associées aux items café (C) et thé vert (TV) sont : (C => TV) et (TV => C). Tout d’abord, on constate que le lift est de 0.34, ce qui laisse croire à une association négative, i.e., que les items sont moins souvent achetés ensemble que s’ils étaient indépendants. En effet, on constate que la confiance de (C => TV) ≈ 2,34% et que le support de TV ≈ 6,85%. Ainsi, l’achat de café fait chuter la probabilité d’acheter du thé vert de 6,85% à 2,34% (facteur d’environ 1/3). Il en est de même pour la règle inverse puisque l’achat de thé vert fait chuter la probabilité d’acheter du café de 11,34% à 3,87%. Ce résultat semble plus particulièrement logique pour des transactions impliquant une seule personne puisqu’il est plutôt rare qu’une même personne prendre simultanément un café et un thé vert. Intuitivement, on peut penser que les cas où les deux items seraient achetés simultanément pourraient être des achats pour plus d’une personne.
3. On regarde alors la ligne associée à capuccino (antécédent) dans la matrice de confiance. On devrait donc suggérer une tarte au chocolat (confiance la plus élevée). La probabilité d’acheter une tarte au chocolat sera alors de 47,41%. Sans savoir que le client a acheté un capuccino, la probabilité d’acheter une tarte au chocolat est de 8,42% (support individuel de la tarte au chocolat). On constate que suite à l’achat de capuccino, la probabilité d’acheter de la tarte au chocolat a été multipliée par un facteur d’environ 5,6 (lift de cette paire d’items).