



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



UNIVERSITÉ
CAEN
NORMANDIE

TRAVAUX PRATIQUES INTRODUCTION SCIENCE DE DONNÉES

FCA et treillis de concepts

Noms et Prénoms de l'étudiant :

OROU-GUIDOU Amirath Fara

Numéro de l'étudiants :

22012235

Parcours : M1 IA Science de Données et Santé

Groupe : 1

Sous la supervision :

REYNAUD Justine

Octobre 2023

Table des matières

1	Introduction	2
2	Acquisition et préparation des données	2
2.1	Question 1	2
2.2	Question 2	2
2.3	Question 3	2
3	Contexte formel	2
3.1	Question 1	2
3.2	Question 2	3
3.3	Question 3	3
4	Treillis de concepts	3
4.1	Question 1	3
4.2	Question 2	3
4.3	Question 3	4
4.4	Question 4	4
5	Règles d'associations	5
5.1	Identifiez les motifs fréquents fermés ainsi que les règles d'associations fréquentes	5
5.2	Les motifs fréquents fermés	5
5.3	Les règles d'associations	6
5.4	Discussion	7

1 Introduction

2 Acquisition et préparation des données

2.1 Question 1

En ligne, trouver une pizzeria qui propose une carte en texte ou en PDF 1. Ce sera votre “base de données” initiale.*

Réponse : La carte pour la pizzeria se trouve dans le fichier "menupizza.csv".

2.2 Question 2

Récupérer le texte de la carte et ne conserver que le texte concernant les pizzas et leurs ingrédients.

Réponse : Voir fichier menupizza.csv.

2.3 Question 3

Construire le contexte formel (G, M, I) avec G l'ensemble des pizzas et M l'ensemble des ingrédients. Ce contexte sera écrit au format .rcf.

Réponse :

Construction du contexte formel (G, M, I) avec :

- G : les éléments de G sont des objets et ces objets ici sont les pizzas
- M : les éléments de M sont des attributs et ses attributs ici sont des ingrédients
- I : une relation binaire ($I \subseteq G \times M$)

Voir la construction du contexte formel dans le fichier "menu.py" et le contexte formel dans le fichier "menu.rcf".

3 Contexte formel

Avant de nous intéresser plus en avant sur le treillis de concept, nous nous intéressons aux données à disposition.

3.1 Question 1

Indiquer le nombre de pizzas et d'ingrédients présents sur la carte.

Commande : On entre dans le répertoire : `cd coron-0.8` et après on tape ma commande `./core01_coron.sh ../menu.rcf 2 -names -alg :apriori`

Réponse : La sortie de la commande est dans apriori.txt

- Nombre de pizzas : 31
- Nombre d'ingrédients : 55

3.2 Question 2

Quel est l'ingrédient le moins utilisé ? Dans combien de pizzas est-il utilisé ?

Commande : `./core01_coron.sh ../menu 2 -names -alg :aprioriinverse`

Réponse : La sortie de la commande est dans `aprioriinverse.txt`

J'ai utilisé cette commande parce que ça permet de trouver l'ensemble des objets rares. Et si un objet est rare c'est que tout ses sous-ensembles sont rares. Il y a plusieurs ingrédients moins utilisés mais j'en ai pris qu'un seul.

- Ingrédient le moins utilisé : "Mozzarella di Bufala"
- Le nombre de pizza dans laquelle il est utilisé : Il est utilisé dans 1 Pizza nommé "Margherita di Bufala".

3.3 Question 3

Quel est l'ingrédient le plus utilisé ? Dans combien de pizzas est-il utilisé ?

Commande : J'ai utilisé la commande de apriori pour visualiser l'ingrédient le plus utilisé.

Réponse :

- Ingrédient le plus utilisé : "Mozzarella"
- Le nombre de pizza dans laquelle il est utilisé : "28 pizzas".

4 Treillis de concepts

Nous allons dans un premier temps travailler que sur un sous-ensemble des données.

4.1 Question 1

À l'aide du pré-traitement Filter-DB sur Coron, créer un contexte formel qui ne conserve que les pizzas au saumon. Ici j'ai pris les pizzas ne contenant que des jambons.

Commande : `/pre03_filterDb.sh ../menu.rcf -attributes ../numAttjambon` avec `numAttjambon` le numéro de l'attribut jambon.

Réponse :

Voir l'affichage du contexte formel dans le fichier "pizzaJambon.rcf".

4.2 Question 2

Extraire les motifs fréquents ($\geq 75\%$). Quels sont les ingrédients fréquemment sur la même pizza ?

Commande : `./core01_coron.sh ../pizzaJambon.rcf 75% -names -alg :apriori`

- Extraction des motifs fréquents : voir le fichier "aprioPizzaJambon.txt". On a $FI^1 = 3$
- Les ingrédients fréquemment sur la même pizza
 - {Mozzarella}
 - {Jambon}
 - {Mozzarella, Jambon}

4.3 Question 3

Y a-t-il des ingrédients qui ne sont jamais présents sans un autre ? Si oui lesquels ? Sinon, y a-t-il des ingrédients pour lesquels c'est presque le cas ?

Commande : `/core01_coron.sh ../pizzaJambon.rcf 1 -names -alg :charm`

J'ai utilisé cette commande parce que ça me permet d'afficher les ensembles d'éléments fermés fréquents. Ça me montre l'ensembles des éléments qui ne sont pas présent sans l'autre.

Réponse :

Voir le fichier "CharmPizzaJambon1.txt". La réponse est oui.

Justification

- Il n'y a jamais de jambon sans mozzarella, donc on a :
 - Mozzarella
 - Jambon

4.4 Question 4

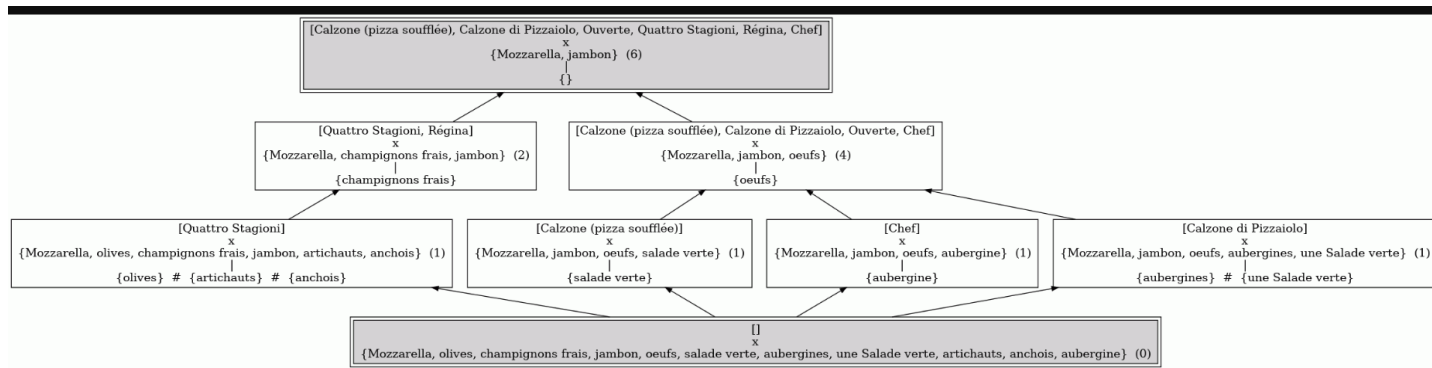
Dessiner le treillis (générer le .dot avec Coron), en affichant les générateurs et les fermés.

Réponse :

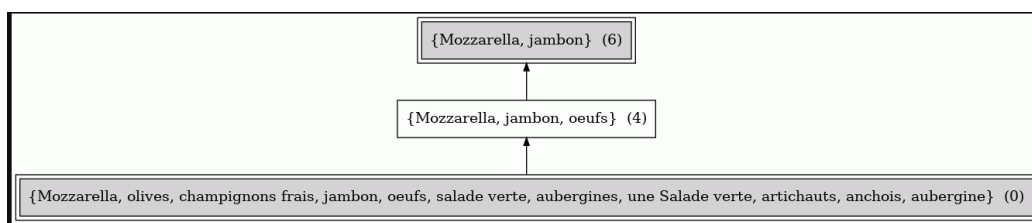
Commande : `/core03_leco.sh ../pizzaJambon.rcf 1 -names -order -alg :dtouch -method :snow -dot` La génération du .dot du treillis est dans le fichier **graphviz/lattice.dot**. Vous avez ci-dessus l'affichage du treillis :

Commande : On entre dans le dossier graphviz : `cd graphviz` et on tape la commande `./compile_gif_leco.sh` et pour voir le treillis on tape la commande : `./view_leco.sh` selon Coron mais ça ne fonctionne pas dans la machine. Il faut aller dans le dossier **graphviz** et cliquer sur le fichier **lattice.gif** [Treillis de Concept](#)

1. ensembles d'éléments fréquents



Treillis Iceberg



5 Règles d'associations

Nous nous intéressons maintenant à l'ensemble du jeu de données. Identifiez les motifs fréquents fermés ainsi que les règles d'associations fréquentes 3 . C'est à vous de décider le seuil à partir duquel un motif ou une règle sont fréquents ; essayez avec 3 valeurs différentes. Dans le comptereendu, vous devrez préciser les valeurs que vous avez choisies et les variations observées selon ces valeurs. Indiquez notamment le nombre de règles obtenues, le nombre d'éléments dans l'antécédent et dans la conséquence, ainsi que la confiance (indiquée par Coron).

5.1 Identifiez les motifs fréquents fermés ainsi que les règles d'associations fréquentes

5.2 Les motifs fréquents fermés

Réponse :

Par rapport à l'affichage des FCIs, j'ai pris différents supports minimums {1, 2, 3}

Commande 1 : `/core01_coron.sh ../menu.rcf 1 -names -alg :charm ;` résultats dans le fichier "**menuFCI1.txt**".

- Minimum support = 1
- $\text{FCI}^2 = 58$

2. ensembles d'éléments fermés fréquents

Commande 2 : `/core01_coron.sh ../menu.rcf 2 -names -alg :charm` ; résultats dans le fichier "**menuFCI2.txt**".

- Minimum support = 2
- **FCI** = 31

Commande 3 : `/core01_coron.sh ../menu.rcf 3 -names -alg :charm` ; résultats dans le fichier "**menuFCI3.txt**".

- Minimum support = 3
- **FCI** = 16

Conclusion : En somme on remarque qu'au fur et à mesure qu'on augmente le support minimal le nombre de motifs fréquents fermés diminue.

5.3 Les règles d'associations

J'ai utilisé l'algorithme de Charm car c'est un algorithme en profondeur très efficace. Ils est plus efficace sur des ensembles de données denses.

- **Commande 1 :** `./core02_assrulex.sh ../menu.rcf 4 60% -names -alg :charm -rule :closed`
- Résultat dans le fichier **regleAsso2.txt**
 - Minimum support : 4
 - Minimum confiance 60%
 - Nombre de règles obtenues : 13
 - Nombre de règles fréquents : 13
 - Le nombre d'éléments dans l'antécédent (ce qui précède la flèche \Rightarrow ³) varie d'une règle à l'autre, allant d'un seul élément à des combinaisons plus complexes d'éléments.
 - Le nombre d'éléments dans la conséquence (ce qui suit la flèche \Rightarrow ⁴) varie d'une règle à l'autre, allant aussi d'un seul élément à des combinaisons d'éléments.
 - La confiance, exprimée en pourcentage, est un indicateur de la fiabilité de la règle. Dans les règles présentées, la confiance est principalement de 100%, indiquant une forte probabilité que la conséquence se produise lorsque l'antécédent est vrai. Cependant, certaines règles affichent une confiance légèrement inférieure, à 66.67%, ce qui signifie que la conséquence a une probabilité de 66.67% de se produire lorsque l'antécédent est vérifié.
- **Commande 2 :** `./core02_assrulex.sh ../menu.rcf 5 70% -names -alg :charm -rule :closed`
- Résultat dans le fichier **regleAsso3.txt**
 - Minimum support : 5

3. suppL

4. suppR

- Minimum confiance 70%
- Nombre de règles obtenues : 6
- Nombre de règles fréquents : 6
- Les règles ont un antécédent simple, comprenant un seul élément, à savoir l'ingrédient spécifique (par exemple, "crème fraîche", "roquette", "poivrons", "oignons", "jambon" et "oeufs").
- La conséquence de toutes les règles est "Mozzarella".
- La confiance dans ces règles est de 100% dans la plupart des cas, sauf pour une règle qui présente une confiance de 83.33%. Cela signifie qu'il y a une forte probabilité que la conséquence (Mozzarella) se produise lorsque l'antécédent est vrai.
- **Commande 3 :** `./core02_assrutex.sh ../menu.rcf 6 80% -names -alg :charm -rule :closed`
- Résultat dans le fichier **regleAsso4.txt**
 - Minimum support : 6
 - Minimum confiance 80%
 - Nombre de règles obtenues : 3
 - Nombre de règles fréquents : 3
 - Les règles ont un antécédent simple, comprenant un seul élément, soit "oignons", "jambon," ou "oeufs."
 - La conséquence de toutes les règles est "Mozzarella".
 - La confiance dans ces règles est de 100%, ce qui signifie qu'il y a une forte probabilité que la conséquence (Mozzarella) se produise lorsque l'antécédent est vérifié.

Conclusion : En somme on remarque qu'au fur et à mesure qu'on augmente le support minimal et la confiance minimale, le nombre de règle diminue.

5.4 Discussion

On remarque que :

- Les trois ensembles de règles montrent comment l'ajustement des paramètres de support et de confiance peut influencer les résultats de l'analyse des règles d'association.
- Un support et une confiance plus élevés conduisent à des règles plus restrictives mais plus fiables, tandis qu'un support et une confiance plus bas permettent de découvrir un plus grand nombre d'associations potentielles.
- Le choix des paramètres dépend des objectifs de l'analyse. Si vous cherchez des associations spécifiques et fiables, des paramètres plus stricts sont appropriés. Si on cherche à explorer un large éventail d'associations, des paramètres moins restrictifs sont préférables.
- Ces règles d'association sont utiles pour découvrir des tendances et des liens entre les ingrédients du menu de la base de données. La Mozzarella est un ingrédient clé qui est souvent associé à d'autres, ce qui peut être utile pour la création de nouvelles recettes ou la prise de décisions en matière de menu.