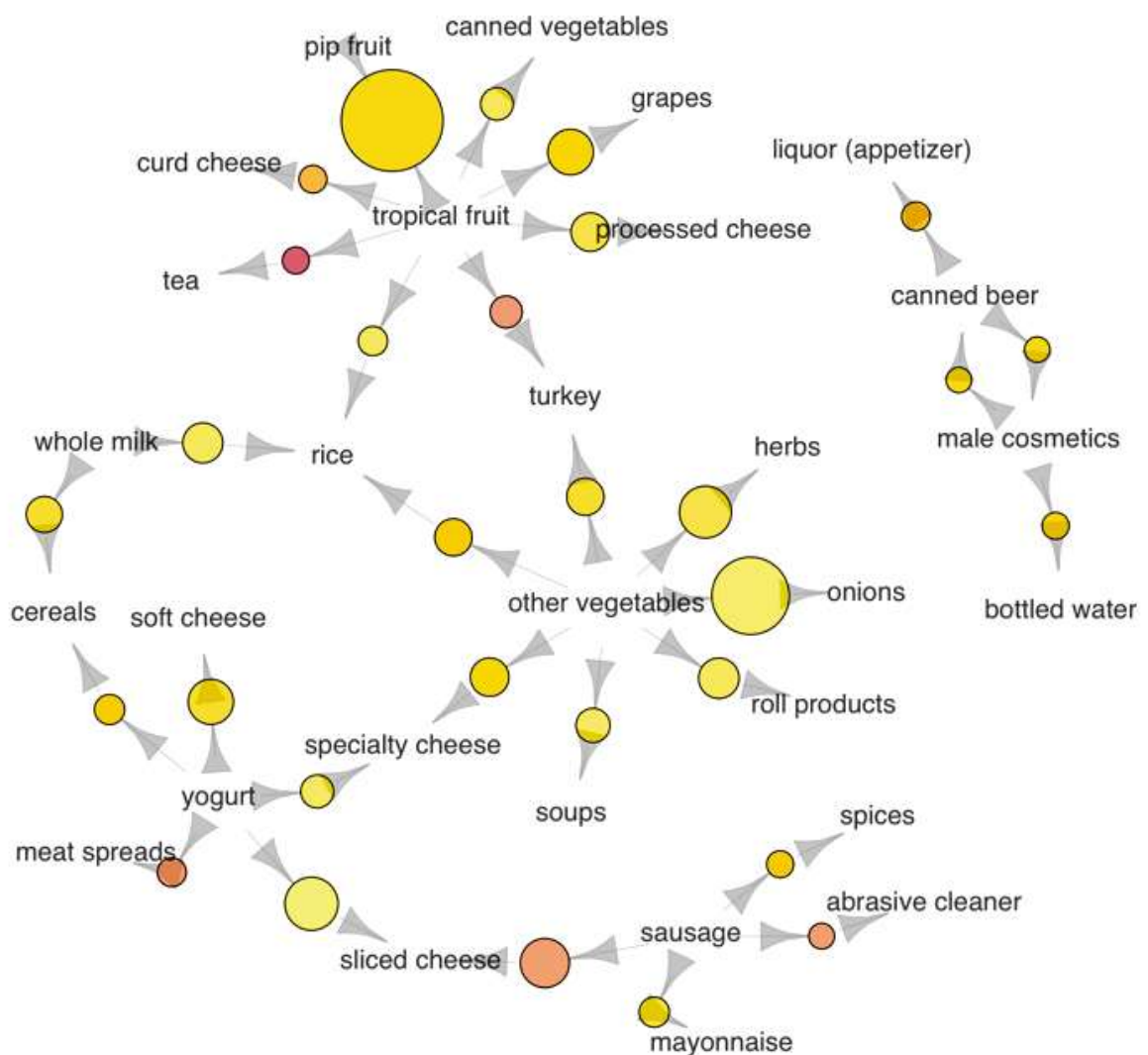


Apprentissage non-supervisé :

Principe Algorithme Apriori

L'algorithme d'Apriori fait référence à l'algorithme qui est utilisé pour calculer les règles d'association entre les objets. La règle d'association décrit comment deux ou plusieurs objets sont liés les uns aux autres. L'algorithme d'Apriori est également appelé "frequent pattern mining".

En général, l'algorithme d'Apriori est utilisé sur une base de données qui comprend un grand nombre de transactions. L'intérêt de cet algorithme est d'améliorer l'intérêt de l'utilisateur grâce à l'identification de l'importance des ensembles d'objets et de leurs occurrences fréquentes.



Quels sont les éléments de l'algorithme ?

L'algorithme d'Apriori comporte trois composantes principales qui sont les suivantes : le support, la confiance et la levée.

-Le support :

Le support est la popularité par défaut d'un produit(A). Le support est le quotient de la division du nombre de transactions comprenant ce produit A par le nombre total de transactions.

$$\text{Support (A)} = \frac{\text{Number of transaction in which A appears}}{\text{Total number of transactions}}$$

-La confiance :

Elle fait référence à la possibilité que les clients aient acheté à la fois deux produits A et B. Vous devez donc diviser le nombre de transactions comprenant à la fois le produit A et le produit B par le nombre total de transactions pour obtenir la confiance.

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{Support(A} \cup \text{B)}}{\text{Support(A)}}$$

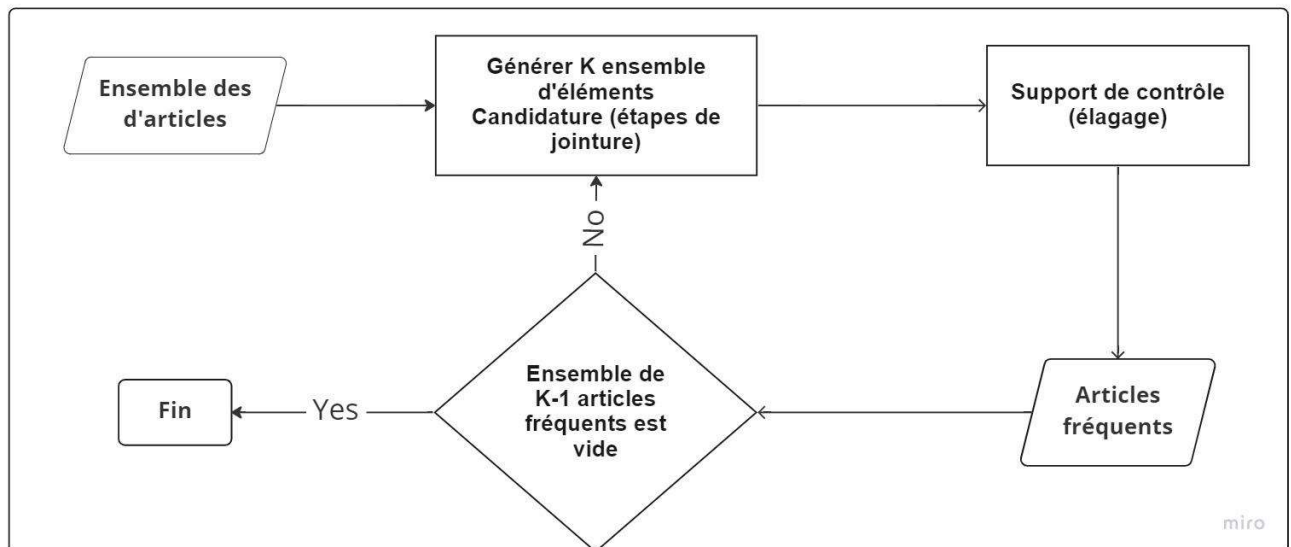
-La levée :

Elle désigne l'augmentation du ratio de la vente d'un article X lorsque vous vendez un autre article Y.

$$\begin{array}{l} \text{Rule } X \Rightarrow Y \begin{cases} \text{Support} = \frac{\text{Frequency}(X,Y)}{N} \\ \text{Confidence} = \frac{\text{Frequency}(X,Y)}{\text{Frequency}(X)} \\ \text{Lift} = \frac{\text{Support}}{\text{Support}(X) * \text{Support}(Y)} \end{cases} \end{array}$$

Comment fonctionne l'algorithme Apriori ?

L'algorithme d'Apriori est une séquence d'étapes à suivre pour trouver les éléments les plus fréquents dans une base de données. Un seuil de support minimum est supposé par l'utilisateur.



1) Dans la première itération de l'algorithme, chaque élément est considéré comme un candidat "1-itemsets". L'algorithme va compter les occurrences de chaque élément. C'est ce qu'on appelle la génération de l'ensemble de candidats.

2) L'ensemble des 1 - itemsets dont l'occurrence satisfait le seuil de support minimum (min_sup) est déterminé. Seuls les candidats dont le nombre d'occurrences est supérieur ou égal à min_sup sont retenus pour l'itération suivante. Les autres sont considérés comme peu fréquents et sont donc supprimés. Cette étape d'élagage est réalisée pour réduire la taille des ensembles d'éléments candidats.

3) Ensuite, nous allons passer à l'étape de jonction. Cette étape génère (K+1) jeu d'articles à partir de K-jeux d'articles en joignant chaque item avec lui-même. Les éléments fréquents à deux ensembles avec min_sup sont découverts.

4) Les candidats au 2-itemset sont élagués en utilisant la valeur seuil min_sup. Maintenant la table aura 2 -itemsets.

5) L'itération suivante formera 3-itemsets en utilisant l'étape de jointure et d'élagage. Cette itération suivra la propriété anti monotone (les sous-ensembles de 3-itemsets, c'est-à-dire les sous-ensembles de 2-itemsets de chaque groupe tombent dans min_sup. Si tous les sous-ensembles de 2-itemsets sont fréquents, alors le sur-ensemble sera fréquent, sinon il sera élagué).

6) L'étape suivante consistera à créer le 4-itemset en joignant le 3-itemset à lui-même et en élaguant si son sous-ensemble ne répond pas au critère min_sup. L'algorithme est arrêté lorsque l'ensemble le plus fréquent est atteint.

Références :

<https://www.javatpoint.com/apriori-algorithm>

<https://www.softwaretestinghelp.com/apriori-algorithm/>

<https://www.engati.com/glossary/apriori-algorithm>