

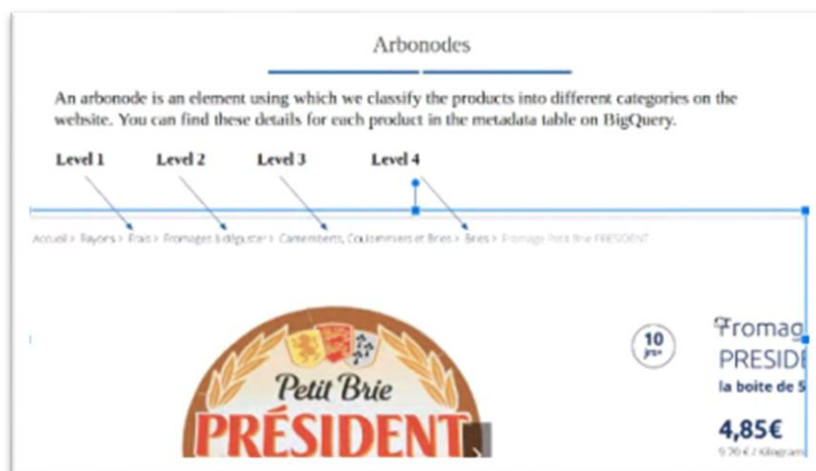
Conférence avec Carrefour

Commencer avec le fichier part5.zip afin de ne pas avoir trop de données à manipuler. Voir s'il faut encore récupérer le fichier de métadonnées ou bien si on l'a.

Quel est le but du sujet

Imaginons que l'on souhaite prédire produit par produit. On peut par exemple faire cela à partir du code-barre. Mais cela risque d'être compliqué et ça n'est pas l'idée de projet. Il y a des API qui existent mais ça n'est pas l'objectif. Par exemple, prenons le coca 1.5L et 1L, on a des codes-barres différents alors qu'on a des produits similaires.

Plutôt que faire une tâche de classification, l'idée est de prédire la catégorie de produit. Cela permet de proposer les produits similaires. On peut choisir un niveau précis entre 1 et 4, on a qu'une seule information par produit. On peut donc soit choisir une catégorie, soit faire un modèle à plusieurs niveaux. Prendre le *is_primary_link* comme label pour prioriser. Un produit bio peut-être présent dans plus catégorie : un camembert Président peut être à la fois dans la catégorie fromage et la catégorie bio. Le fait de regrouper les données par catégorie augmente la taille de notre ensemble de données.



Pour évaluer notre modèle, il s'agit de séparer notre modèle en ensemble de train/validation et test. On peut tenter de prendre des photos nous-mêmes afin de tester notre modèle pour évaluer sa qualité. Ce qui peut être également intéressant c'est de voir si un produit d'une autre marque fonctionne dans notre modèle.

Info : OpenFoodFact - <https://fr.openfoodfacts.org/>. Base de données assez riche.

Pour résumer : le plus simple est de faire algorithme de classification au niveau d'un niveau de catégorie. Envoyer un e-mail si on a des questions.