



Classification de produits e-commerce Rakuten

Cursus concerné : Data Scientist

Niveau de difficulté : 08/10

Description du projet :

Cataloguer les produits selon des données différentes (textes et images) est important pour les e-commerces puisque cela permet de réaliser des applications diverses telles que la recommandation de produits et la recherche personnalisée. Il s'agit alors de prédire le code type des produits sachant des données textuelles (désignation et description des produits) ainsi que des données images (image du produit).

Ressources à consulter :

- Données :

- Ce projet s'inscrit dans le challenge **Rakuten France Multimodal Product Data Classification**, les données et leur description sont disponibles à l'adresse : <https://challengedata.ens.fr/challenges/35>
- Les données textuelles : **~60 mb**
- Données images : **~2.2 gb**
- **99k** données avec plus de **1000** classes.

Conditions de validation du projet :

- un **rapport** d'exploration, de data visualisation et de pre-processing des données ;
- un **rapport** de modélisation ;
- un **rapport** final et le **GitHub** associé.