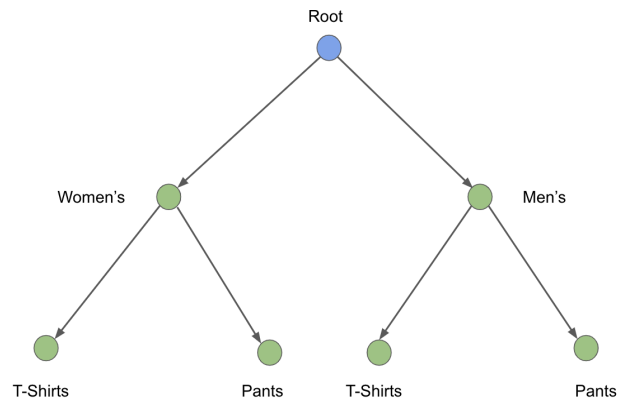


Classification de produits

Sur un site de e-commerce il est important que les bons produits soient placés dans les bonnes catégories. Lorsque cela n'est pas le cas, ils peuvent être invisibles aux acheteurs, mais peuvent également être à l'origine d'une expérience utilisateur désagréable, en particulier lorsqu'ils sont présents en grande quantité dans une catégorie erronée.

Chaque site de e-commerce a sa propre taxonomie de catégories. Voici un exemple:



Un produit peut être décrit par des informations diverses et variées. Dans le cadre de cette étude, les variables explicatives ont d'ores et déjà été transformées et réduites à une dimension de 128.

L'objectif de cette étude est de créer un algorithme de classification sur les catégories feuilles (*category_id*) à partir de ces 128 features (*f0*, ... *f127*).

Les données fournies:

- ***data_train.csv***: contient les colonnes suivantes:
 - *product_id*: ids des produits
 - *category_id*: catégories feuilles des produits
 - *f0* ... *f127*: les 128 features
- ***data_test.csv***: même format que *data_train.csv*
- ***category_parent.csv***: contient les paires d'ids, chacune correspondant à une catégorie et son parent, -1 étant le Root.

Remarque: Etant donné que des transformations ont d'ores et déjà été effectuées sur les données, les variables ne sont pas interprétables.

Guidelines:

- Analyser les données et décrire brièvement quelques algorithmes de classification possibles sur un tel dataset.
- Choisir et implémenter au minimum un de ces algorithmes en utilisant le set du training (*data_train.csv*) et effectuer une prédiction sur le set du test fourni (*data_test.csv*). Préciser les métriques choisies. Expliquer vos démarches.
- Effectuer une analyse des résultats obtenus.
- Ne pas utiliser la hiérarchie des catégories comme variable explicative.

Pour le format de l'étude, un notebook est préférable (github, envoi par email, sites de partage de fichiers, etc.) et doit être reproductible. Joignez un *requirement.txt* en précisant la version des librairies si cela vous semble nécessaire.

Bon courage!