

## Contexte

En France, les prix des médicaments sont pour la plupart administrés, même si des prix libres existent pour certains d'entre eux. Le processus de fixation des prix suit un cadre réglementaire bien précis [1] [2].

Pourtant, des écarts de prix peuvent subsister [3].

L'objectif de ce challenge est de mettre au point un modèle prédictif pour estimer combien coûte une boîte de médicaments, afin par exemple d'aider à en fixer le prix, ou bien détecter des anomalies dans les tarifs.

## Critère de performance

La métrique utilisée pour mesurer la précision du modèle est MAPE : Mean Absolute Percentage Error : [https://en.wikipedia.org/wiki/Mean\\_absolute\\_percentage\\_error](https://en.wikipedia.org/wiki/Mean_absolute_percentage_error)

## Données

Un même médicament peut se commercialiser sous différentes formes (taille de la boîte, type de gélule, ...) faisant varier le prix de la boîte.

Le fichier principal à utiliser pour le concours est **boites\_medicaments.csv**.

Il est encodé en UTF-8, le séparateur est « ; ».

Il contient les informations de chaque boîte de médicament, et leur prix :

**libelle** : description du contenu de la boîte de médicament (exemple : 5 ampoules en verre de 5 ml)

**libelle\_xxx** : 13 variables qui codent la présence de certains mots dans le libellé (exemple libelle\_ampoule = 1 si le mot « ampoule » apparaît dans le champs « libelle », 0 sinon.

**nb\_xxx** : 14 variables qui indiquent certaines quantités présentes dans les libellés (exemple : nb\_ampoule = 32 si « 32 ampoules » apparaît dans le champ « libelle », sinon 0

**statut** : statut administratif de la boîte de médicament

**etat commerc** : Etat de commercialisation de la boîte de médicament tel que déclaré par le titulaire de l'AMM

**agrement col** : agrément aux collectivités ("oui", "non" ou « inconnu »)

**tx rembours** : taux de remboursement

**forme pharma** : forme pharmaceutique (comprimé pelliculé, comprimé sécable, ...)

**voies admin** : voies d'administration (avec un séparateur « , » entre chaque valeur quand il y en a plusieurs : orale, ophtalmique, intramusculaire, ...)

**statut admin** : statut administratif de l'autorisation de mise sur le marché (AMM) : active, abrogée,

...

**date declar année** : année de la déclaration de commercialisation de la boîte de médicament

**date amm année** : année d'Autorisation de Mise sur le Marché du médicament

**type proc** : Type de procédure d'autorisation de mise sur le marché (AMM) (Nationale, décentralisée, reconnaissance mutuelle, ...)

**titulaires** : laboratoire titulaire du médicament. S'il y a plusieurs titulaires, ils sont séparés par des « , »

**substances** : substances présentes dans le médicament. S'il y en a plusieurs, elles sont séparées par des « , »

**prix** : le prix de la boîte de médicament (présent dans le fichier « train » uniquement, à prédire dans le fichier « test »)

## **Format du fichier de prédictions**

Le format du fichier à soumettre est le suivant :

id;prix

1;2.3987

2;7.3873

3;6.9833

[1] <http://www.irdes.fr/documentation/syntheses/historique-de-la-politique-du-medicament-en-france.pdf>

[2] <http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:31989L0105>

[3] <http://www.irdes.fr/Publications/2012/Qes178.pdf>