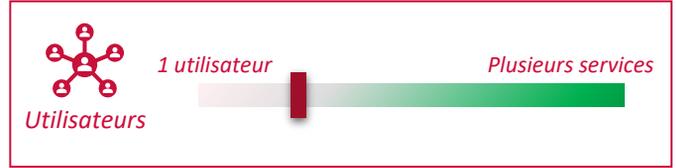
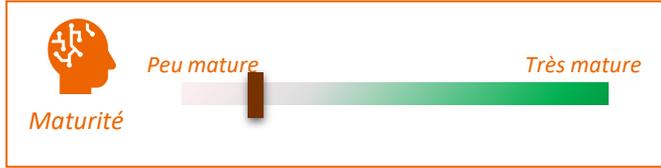


Objectif : Construire une démarche de structuration des données cliniques à partir de l'interprétation des comptes rendus médicaux par des approches d'IA afin d'alimenter l'entrepôt de données institutionnel nécessaire aux travaux de recherche en cancérologie.

Utilisateurs cibles : Chercheurs

Dates clefs et avancement : Projet lancé en 2019 – En phase de recherche / développement



Données utilisées : Comptes rendus médicaux
Hébergement : Local

Recherche médicale
Traitement du langage naturel
Mots clés
Codage des données de santé

Nature de la solution

- > Classification de **données textuelles** par analyse de mots-clefs et association d'informations par **item** (catégorie de thématique de recherche)
- > Pour chaque item clinique :
 - Apprentissage à partir de **phrases annotées manuellement** issues du cluster d'intérêt
 - Entraînement sur un set d'algorithmes ayant démontré leur potentiel
 - **Sélection du meilleur algorithme**

Modalité d'évaluation et de validation

- > Vérification de la qualité des extractions par **comparaison entre** :
 - **base de données structurées** générées par la solution
 - d'autres **bases de données de référence** (registres, bases institutionnelles)
- > Modèle fiable si le **score de performance est supérieur à 90-95%**

Origine du projet

- > Travaux d'annotation des données chronophages manuellement
- > Depuis 2014 : ce travail ne peut plus être réalisé de façon exhaustive au sein de l'Institut Curie au regard des volumes de données

Modèle de diffusion

- > Accompagnement à l'appropriation de l'outil par les équipes de l'Institut
- > Conduite de chaque projet de structuration d'un item clinique par un **binôme médecin/data scientist** pour garantir la qualité des résultats et éviter les projets doublons

Perspectives

- > Généralisation de l'usage de l'outil à d'autres items cliniques
- > Mise en Open Source du code source de l'outil pour permettre l'usage à d'autres établissements

Facteurs clés de succès

- ✓ Transparence du processus et explicabilité des résultats obtenus
- ✓ Implication de professionnels de santé experts dans le projet
- ✓ Tests de plusieurs types d'algorithmes pour identifier la solution la plus adaptée

Organisation du projet

- Gestion de projet :** équipe interne à l'Institut (ingénieurs, datamanagers, Direction des données et Conseil médical et Stratégique de l'Institut)
- Partenaires :** recherche d'un partenaire universitaire concernant la partie techniques de développement de l'IA
- Financements :** interne à l'Institut

Contacts

- Julien Guerin, responsable de l'équipe « Data Factory » à la Direction des Données, julien.guerin@curie.fr
- Hijano Cubelos Oliver – ingénieur IA et porteur du projet
- Thomas Balezeau (ingénieur en bioinformatique) et Armand Leopold (responsable infrastructure et sécurité).
- Dr. Alain Livartowski – Conseiller médical et stratégique