

Stage de Master Informatique Data Science Hôpital Nord Franche-Comté Assistance au codage de séances de radiothérapie par IA.

Philippe Selles¹ and David Laiymani, Jean-François Couchot, Christophe Guyeux²

¹Hôpital Nord Franche-Comté: philippe.selles@hnfc.fr

²Institut FEMTO-ST: couchot@femto-st.fr

3 décembre 2024

1 Contexte

Localisé à Trévenans entre Belfort et Montbéliard, l'Hôpital Nord Franche-Comté (HNFC) est celui de l'aire urbaine de Belfort-Montbéliard-Héricourt dans le nord de la Franche-Comté. Doté de 1200 lits, ce sont 100000 séjours de patients qui sont réalisés approximativement chaque année.

Cet hôpital possède par ailleurs un service de radiothérapie réalisant approximativement 50000 séances par an. Chacune de ces séances est préparée en amont de la première séance d'irradiation du patient grâce à un outil logiciel commercial utilisant de l'Intelligence Artificielle (IA) afin de maximiser l'irradiation du volume cible (tumeur cancéreuse) et minimiser la dose de rayonnement ionisant aux organes sains. Mensuellement et après les séances, des codes diagnostics (CIM10) et actes (CCAM) sont associés à ces séances et avec la seule intelligence humaine des codeurs et des codeuses pour des besoins de facturation des séances de radiothérapie réalisées à l'Assurance Maladie.

L'objectif de ce stage est de faciliter la démarche d'association de codes en exploitant l'IA.

2 Associer automatiquement des codes : un savoir faire à concrétiser

Comme il s'agit d'associer plusieurs ensembles de codes à un ensemble de données, il s'agit d'un problème de classification multilibellés.

L'HNFC a déjà une bonne expérience pour cette famille de problèmes, puisqu'il a déjà développé un classifieur permettant d'attribuer automatiquement des codes CIM-10¹ à des parcours de soins décrits au moyen de comptes-rendus. Ce travail a été réalisé en partenariat avec des membres du laboratoire FEMTO-ST, participant à l'encadrement de ce stage. Les contributions scientifiques ont été valorisées par des articles scientifiques [3, 1, 2]. Un outil opérationnel a été développé et a obtenu des résultats remarquables. Tous les codes Python sont accessibles en ligne².

Par rapport aux réalisations précédentes, le travail n'est plus le même pour différentes raisons. D'une part les données sont différentes. Précédemment, le dataset était constitué des documents textuels et des codes CIM-10 associés. Ici les données vont comprendre des tableaux (paramètres de radiothérapie), des textes et les codes associés. D'autre part, les codes seront différents. Il s'agira d'une part des codes diagnostics CIM10 mais avec une moins grande variabilité que dans les travaux précédents : essentiellement les codes parfaitement définis par convention pour le diagnostic principal. D'autre part, d'autres codes (CCAM) caractérisant la préparation et le traitement de radiothérapie seront à générer.

Les modèles d'IA de classification multilibellés ne pourront donc pas être directement inspirés de ceux des travaux précédents.

1. https://fr.wikipedia.org/wiki/Liste_de_codes_CIM-10

2. <https://github.com/mlfiab>

Une interface simple devra enfin être construite, permettant d'automatiser au maximum cette tâche.

3 Profile type

La personne recrutée doit avoir une réelle envie de parfaire ses connaissances en intelligence artificielle et de donner du sens au travail réalisé. Il s'agit ici de travailler en IA dans le domaine de la santé, avec un impact direct pour l'Hôpital Nord Franche-Comté.

4 Organisation

L'ensemble des données traitées sont strictement couvertes par le secret médical. L'accès à celles-ci, leur utilisation et exploitation se fera sous la responsabilité du chef de service du Département de l'Information Médicale (DIM). Cela implique une confidentialité totale au vu de la sensibilité des informations et un engagement écrit à respecter la charte informatique de l'HNFC. Tout manquement caractérisé fera l'objet d'un dépôt de plainte auprès de la juridiction compétente.

Le stage se déroulera au DIM de l'HNFC, au cœur de l'aire urbaine de Belfort-Montbéliard-Héricourt. L'Hôpital est desservi par une ligne régulière de bus (toutes les 15 minutes en moyenne) depuis Belfort et Montbéliard et à 2 km de la gare TGV de Belfort-Montbéliard.

Le stage durera entre 4 et 6 mois, à partir de janvier 2025 suivant les possibilités de la personne recrutée. Il donnera lieu à une gratification habituelle.

5 Contacts

Les personnes intéressées peuvent prendre contact directement avec les responsables du stage.

Références

- [1] Yakini Tchouka, Jean-François Couchot, and David Laiymani. An easy-to-use and robust approach for the differentially private de-identification of clinical textual documents. In Federico Cabitza, Ana Fred, and Hugo Gamboa, editors, *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2023, Volume 5 : HEALTHINF, Lisbon, Portugal, February 16-18, 2023*, pages 94–104. SCITEPRESS, 2023.
- [2] Yakini Tchouka, Jean-François Couchot, David Laiymani, Philippe Selles, and Azzedine Rahmani. Automatic ICD-10 code association : A challenging task on french clinical texts. In João Rafael Almeida, Myra Spiliopoulou, José Alberto Benítez-Andrades, Giuseppe Placidi, Alejandro Rodríguez González, Rosa Sicilia, and Bridget Kane, editors, *36th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2023, L'Aquila, Italy, June 22-24, 2023*, pages 91–96. IEEE, 2023.
- [3] Yakini Tchouka, Jean-François Couchot, David Laiymani, Philippe Selles, and Azzedine Rahmani. Differentially private de-identifying textual medical document is compliant with challenging NLP analyses : Example of privacy-preserving ICD-10 code association. *Intell. Syst. Appl.*, 23 :200416, 2024.