

Classification explicable de textes médicaux par apprentissage supervisé

Ce travail présente une chaîne complète de **classification explicable** de textes médicaux destinée à soutenir l'usage clinique de modèles NLP. Nous considérons un corpus d'abrévés biomédicaux annotés en plusieurs catégories et comparons de façon systématique des représentations textuelles (TF-IDF, Word2Vec, GloVe, **SBERT**, **BioBERT**) couplées à des classificateurs supervisés (**Naïve Bayes**, **Arbre de décision**, **SVM linéaire**, **Régression logistique**). L'évaluation repose sur une validation croisée stratifiée à 5 plis et sur des mesures standard (exactitude, précision, rappel, F1 macro), avec contrôle de la robustesse (seeds multiples, bruit lexical) et de l'équilibre entre classes.

L'apport central du projet est l'intégration native d'une **couche XAI** (Explainable AI) couvrant l'**interprétation globale et locale**.

- Côté global, nous exposons (i) les *log-vraisemblances* par classe de Naïve Bayes pour TF-IDF, (ii) les *coefficients* par classe de la Régression logistique (sparse TF-IDF et denses BioBERT/SBERT), (iii) les *chemins de décision* d'arbres entraînés sur des vecteurs Word2Vec/GloVe, et (iv) l'**importance par permutation** afin d'auditer la dépendance du modèle à des dimensions d'embedding.
- Côté local, nous générons des explications *par instance* avec **LIME** sur SVM et LogReg (calibrés lorsque nécessaire) : les pondérations mot-à-mot ou dimension-à-dimension mettent en évidence les éléments qui ont fait basculer la décision. Nous complétons par des **contre-factuels légers** (ajout/suppression de tokens) pour tester la stabilité des prédictions. Toutes les figures XAI (barres de termes NB, arbres, contributions LIME multi-classes) sont produites automatiquement pour l'intégration au rapport.

Sur le plan des performances, les **représentations contextuelles** dominent les approches classiques : **BioBERT + SVM linéaire** obtient les meilleurs scores globaux (F1 macro et exactitude), suivi de près par **SBERT** avec SVM/LogReg. **Naïve Bayes** reste compétitif avec **TF-IDF** grâce à sa parcimonie et fournit des cartes de termes très lisibles. Les **arbres de décision** sont moins performants sur des espaces denses et de grande dimension, mais leur transparence offre un complément utile d'audit. Les explications concordent avec la connaissance clinique (p. ex., termes cardio-vasculaires sur la classe correspondante) et permettent d'identifier des erreurs systématiques (abréviations, polysémie, bruit OCR).

Nous livrons (i) un **benchmark reproductible** mêlant classiques et transformeurs biomédicaux, (ii) un **module XAI unifié** générant des artefacts globaux/locaux exploitables par des experts, et (iii) des **lignes directrices** pour choisir un couple représentation-classifieur selon les contraintes d'explicabilité et de ressources.

Limites. Taille et diversité du jeu d'entraînement, coût calculatoire de SBERT/BioBERT, et fidélité imparfaite des méthodes post-hoc.