

Manuel Utilisateur – HealthPredict AI



HealthPredict AI

1. Introduction

HealthPredict AI est une application interactive permettant :

- d'analyser des rapports d'incidents médicaux (texte, PDF, DOCX, images scannées via OCR),
- de prédire la criticité des pannes (⚠ critique / ✅ non critique),
- de consulter un tableau de bord interactif avec tendances et indicateurs.

L'application est disponible en mode local (via **Streamlit**) et exploite une base SQLite pour l'historique des prédictions.

2. Installation et lancement

1. Installer un éditeur de code comme Visual Studio Code et ensuite télécharger le projet depuis GitHub (<https://github.com/GemimaOndele/Healthpredict-AI>) / Hugging Face avec cette commande **python scripts/download_assets.py**
2. Installer les dépendances :
3. `pip install -r requirements.txt`
4. Activer l'environnement virtuel et lancer l'application :
5. `streamlit run app/healthpredict_app.py`
6. Ouvrir le lien dans le navigateur : <http://localhost:8501>

3. Navigation dans l'application

L'application se compose de **5 onglets principaux** :

Tableau de bord

- Liste filtrable des équipements et incidents.
- Indicateurs clés (total rapports, dernier incident).
- Graphiques :
 - Fréquence mensuelle des incidents,
 - Répartition par type d'événement,
 - Répartition par type d'équipement.

Prédiction (texte libre)

- Saisir une alerte ou description.
- Choisir le modèle (TF-IDF ou CamemBERT).
- Option : traduction automatique (EN→FR).
- Résultat : prédiction critique / non critique + niveau de confiance.
- Export possible des mots-clés contributifs (CSV/Excel).

Documents

- Importer un ou plusieurs fichiers (PDF, DOCX, TXT, CSV, images).
- OCR automatique pour les documents scannés (Tesseract + Poppler).
- Analyse textuelle avec prédiction + extraction des mots-clés.
- Enregistrement des résultats dans l'historique.

Historique & Prévisions

- Consultation des prédictions sauvegardées (SQLite).
- Recherche par mot-clé ou nom de fichier.
- Tendance mensuelle avec projection linéaire sur 3 mois.
- Export CSV/Excel de l'historique.


Évaluation

- Résultats de l'évaluation des modèles (Accuracy, F1, ROC-AUC).
- Graphiques disponibles : courbe ROC, courbe Precision-Recall, matrice de confusion.

4. Bonnes pratiques

- Vérifier que les données saisies sont **anonymisées** avant usage.
- Pour de meilleures performances :
 - privilégier le modèle **CamemBERT** pour le français,
 - utiliser **TF-IDF** pour les tests rapides ou sur de faibles ressources.
- Toujours relancer l'entraînement (train_minimal_tfidf.py ou train_camembert_baseline.py) si de nouvelles données sont ajoutées.

5. Support et limites

- Ce projet est un **démonstrateur académique** : il ne doit pas être utilisé en milieu hospitalier sans validation experte.
- Problèmes fréquents :
 - OCR incomplet sur documents très bruités,
 - déséquilibre des classes (peu de cas critiques réels).
- En cas de souci technique : se référer au dépôt GitHub ou me contacter :
 gemimakerenondelepourou@gmail.com