# ****Projet : Prédiction du Risque de Cancer via une Interface Web Interactive****

## ****Introduction****

Le cancer demeure l'une des principales causes de mortalité mondiale, avec des millions de nouveaux cas diagnostiqués annuellement. Un diagnostic précoce et des mesures préventives adaptées peuvent significativement améliorer les pronostics vitaux. Ce projet propose le développement d'une plateforme web innovante permettant aux utilisateurs d'évaluer leur risque de développer un cancer sur la base de leurs paramètres médicaux et comportementaux.

En associant des algorithmes d'apprentissage automatique à une interface utilisateur intuitive, cette solution offre une évaluation personnalisée du risque oncologique, incitant à une consultation médicale précoce le cas échéant et à l'adoption de comportements préventifs.

## ****Alignement avec les Objectifs de Développement Durable****

Cette initiative s'inscrit pleinement dans le cadre de l'****Objectif de Développement Durable 3 : Bonne santé et bien-être****, en contribuant à :

****Renforcer la médecine préventive**** par l'identification des populations à risque avant la manifestation clinique de la maladie

****Réduire la mortalité cancéreuse**** grâce à une détection précoce permettant des interventions thérapeutiques plus efficaces

****Optimiser l'allocation des ressources sanitaires**** en orientant les campagnes de dépistage vers les populations les plus exposées

****Revue de Littérature Scientifique****

Les recherches actuelles démontrent l'efficacité croissante des modèles prédictifs en oncologie :

L'étude "Machine Learning Techniques for Cancer Prediction" (Journal of Healthcare Engineering) établit une comparaison méthodique des différents algorithmes de classification appliqués à la prédiction cancérologique, mettant en évidence la supériorité des approches supervisées comme les forêts aléatoires dans ce contexte.

Par ailleurs, la publication "Predicting Cancer Risk Using Lifestyle and Medical Data" (International Journal of Medical Informatics) démontre que l'intégration de données comportementales (tabagisme, activité physique, habitudes alimentaires) aux paramètres cliniques améliore significativement la précision des modèles prédictifs.

## ****Analyse et Préparation des Données****

### ****Source des données****

Le jeu de données provient de la ressource "Cancer Prediction Dataset" disponible sur Kaggle (Rabie El Kharoua).

### ****Caractéristiques principales****

Volume : 1 500 observations

Variables :

Paramètres médicaux : antécédents familiaux, comorbidités, marqueurs biologiques

Facteurs comportementaux : consommation de tabac, niveau d'activité physique, habitudes nutritionnelles

### ****Traitement préalable****

Imputation des valeurs manquantes par méthodes appropriéesTransformation des variables catégorielles via encodage adapté

Normalisation des variables numériques pour l'analyse

## ****Modélisation Prédictive****

## ****Sélection des algorithmes****

La nature structurée des données justifie l'emploi d'approches supervisées, avec une évaluation comparative de :

Régression logistique (modèle de référence)

Forêts aléatoires (résistance au sur-apprentissage)

Machines à vecteurs de support (classification non linéaire)

XGBoost (optimisation par boosting)

### ****Évaluation des performances****

Métriques : précision, sensibilité, score F1

Analyse : courbe ROC et aire sous la courbe (AUC)

## ****Architecture Technique****

### ****Fonctionnalités principales****

Saisie standardisée des données utilisateur

Calcul instantané du risque estimé

Visualisation analytique des résultats

Recommandations personnalisées

### ****Environnement technologique****

Interface : framework JavaScript moderne (React)

Backend : microservices Python (Flask)

Modèle : implémentation via scikit-learn

****Impact et Perspectives****

Cette plateforme présente un potentiel significatif pour :

Améliorer la sensibilisation aux facteurs de risque

Favoriser le recours aux consultations préventives

Évoluer vers un système intégré de surveillance sanitaire