WILD
CODE
SCHOOL

**Projet 3 : W-LAB**

WILD CODE SCHOOL

# Health Analytics Project: WILD LAB Analytics (W-LAB)

WILD CODE SCHOOL

# Objectifs du projet :

- Acquérir des connaissances de base en recherche scientifique, y compris la documentation, l'état de l'art, et la compréhension du métier.

- Développer un modèle prédictif capable de classifier les individus en fonction de la présence ou de l'absence des maladies en se basant sur des variables médicales.

- Concevoir une application permettant à l'utilisateur de prédire le risque de développer l'une des maladies suivantes :

    - Diabètes
    - Cancer du sein
    - Maladie rénale chronique
    - Maladie chronique cardiaque
    - Maladie du foie

WILD CODE SCHOOL

# Plus concret :

Développement d’application et déploiement de Modèles de Prédiction pour les Utilisateurs : Intégrer les modèles de machine learning au sein d'une application qui permet aux utilisateurs de saisir leurs propres données médicales et de recevoir des prédictions sur leur état de santé. Cela inclut :

- Le chargement des modèles pré-entraînés pour chaque maladie (diabète, cancer du sein, risque de maladies cardiaques, maladie rénale chronique, maladie du foie).
- La mise en place d'une interface pour la saisie des données utilisateurs selon les caractéristiques requises par chaque modèle.
- La fourniture de résultats de prédiction aux utilisateurs avec un avertissement clair que ces prédictions sont informatives et ne remplacent pas un diagnostic médical professionnel.

WILD CODE SCHOOL

# But pédagogique :

1. Manipulation de Données de Santé : Acquérir des compétences en nettoyage, transformation et analyse préliminaire de données médicales complexes.

2. Classification Binaire : Comprendre et appliquer la classification binaire pour identifier les risques de maladies chroniques (Cancer du sein, Maladies cardiaques, Insuffisance rénale chronique, Maladies du foie.

3. Analyse Multivariée et Sélection de Caractéristiques : Maîtriser l'analyse multivariée pour sélectionner des biomarqueurs clés dans la prédiction de maladies.

WILD CODE SCHOOL

# But pédagogique :

4. Techniques de Machine Learning en Médecine : Apprendre à utiliser des techniques d'apprentissage automatique adaptées aux données de santé.

5. Interprétation Clinique des Analyses : Développer la capacité d'interpréter les résultats d'analyses de données de santé dans un contexte clinique.

6. Éthique et Confidentialité en Santé : Comprendre les enjeux éthiques et la confidentialité liés à l'utilisation des données médicales.

WILD CODE SCHOOL

# Descriptif :

1. Exploration de Données Médicales : Examiner les datasets spécifiques à chaque maladie pour comprendre leurs caractéristiques uniques, y compris les défis tels que les valeurs manquantes, le déséquilibre des classes, et la spécificité des données médicales.

2. Nettoyage et Préparation des Données : Traiter et nettoyer les données médicales pour assurer la qualité et l'intégrité requises pour les analyses. Cela comprend la gestion des valeurs manquantes, la normalisation des données et la transformation des variables catégorielles si nécessaire.

3. Sélection de Caractéristiques : Identifier et sélectionner les caractéristiques les plus pertinentes pour chaque maladie en utilisant des techniques d'analyse multivariée, en tenant compte de l'importance clinique et de la pertinence des biomarqueurs.

WILD CODE SCHOOL

# Descriptif :

4. Analyse Statistique et Descriptive : Réaliser des analyses statistiques pour résumer et comprendre les caractéristiques des données, telles que la distribution, la moyenne, la médiane, et l'écart-type des différentes variables.

5. Visualisation des Données : Créer des visualisations claires et informatives pour révéler des tendances, des schémas, et des corrélations importantes dans les données de santé.

6. Modélisation Prédictive : Construire et entraîner des modèles de classification binaire pour prédire la présence ou l'absence de chaque maladie. Expérimenter avec différents algorithmes de machine learning et évaluer leur performance à l'aide de métriques appropriées.

WILD CODE SCHOOL

# Descriptif :

7. Interprétation et Validation des Résultats : Interpréter les résultats des modèles en termes de signification clinique, et effectuer des validations croisées pour assurer la robustesse et la fiabilité des prédictions.

8. Réflexion sur l'Éthique et la Confidentialité : Mener une réflexion critique sur les enjeux éthiques et les considérations de confidentialité liés à l'analyse des données de santé, en soulignant l'importance du consentement éclairé et de la protection des informations sensibles des patients.

WILD CODE SCHOOL

# Techno

Langages de Programmation
- Python

Bibliothèques et Frameworks
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Scikit-learn
- Joblib

Outils de Nettoyage et de Préparation des Données :
- Techniques de traitement des valeurs manquantes
- Encodage des variables catégorielles
- Normalisation et standardisation des données

Analyse Statistique :
- Tests statistiques
- Calcul de la moyenne, médiane, mode, écart-type

Visualisation des Données :
- Graphiques à barres, histogrammes, boîtes à moustaches
- Cartes de chaleur

Modèles de Machine Learning :
- Régression logistique
- Forêts aléatoires
- Machines à vecteurs de support (SVM)
- Réseaux de neurones
- XGBoost, LightGBM (selon la complexité souhaitée)

Métriques d'Évaluation de Modèle :
- Précision
- Rappel
- F1-Score
- AUC-ROC

# Techno

Techniques de Sélection de Caractéristiques :

- Analyse en composantes principales (PCA)
- Méthodes basées sur l'importance des caractéristiques

Outils de Déploiement :

- Interface utilisateur pour la saisie des données (potentiellement via une interface web ou une application)
- Système de gestion de base de données (si la collecte et le stockage des données utilisateur sont nécessaires)

Considérations Éthiques et Juridiques :

- Connaissance des règlements sur la confidentialité des données (comme le RGPD)
- Principes éthiques dans le traitement des données de santé

Environnement de Développement et Outils Associés :

- Jupyter Notebook ou Lab
- Git pour le versionnage du code
- Environnement virtuel Python (comme virtualenv ou conda)

Liens du projet :

https://docs.google.com/document/d/15BQdXImZB9SXViZK1HNh8iDduNZqpV76VCG0DmDvk8Y/edit?usp=sharing

https://docs.google.com/document/d/1lCgkIayp6YMyYvP3o2Znce5oc7ysufFzqtn8Qs1kd1k/edit?usp=sharing

DataSets : (A actualiser)

Délivrés la semaine 2 du projet