



**BIOTECH**  
INSIGHTS

DATA ANALYSIS

# BIOTECH

## INSIGHTS

2024

MEDICAL PROFESSIONAL APP



# TABLE DES MATIERES

01

## Contexte

Développement d'une application à destination des professionnels de santé.

02

## Méthodologie

Process de développement : EDA & ML

03

## L'Application

Déploiement et présentation de l'application BioTech Insights depuis la plateforme Streamlit.

04

## Evolution

Les perspectives d'évolutions de l'application.



# À NOTRE PROPOS

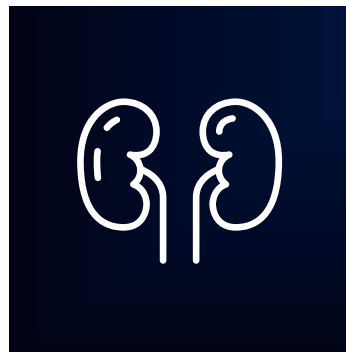
Dans le cadre de notre activité chez BioData Insights, nous avons développé une application médicale innovante.

---

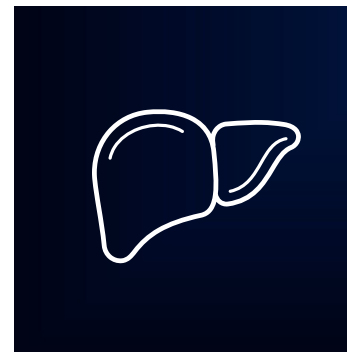
Elle a pour but d'évaluer le risque de développer **cinq maladies majeures** :



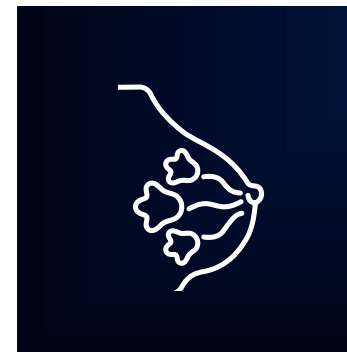
LES MALADIES  
CARDIAQUES



LA MALADIE  
CHRONIQUE  
RÉNALE



LES MALADIES  
DU FOIE



LE CANCER  
DU SEIN



LE DIABÈTE



# NOTRE ÉQUIPE

Rencontrons les membres de notre équipe qui sont professionnels et possèdent une expérience riche.



DOROTHEE CATRY

Data Analyst



GRÉGOIRE LOZAC'H

Data Scientist



CAROLINE ROUVILLAIN

Founder



# TIMELINE

Evolution temporelle du projet

---

Le processus EDA  
(Exploratory Data Analysis)  
permet de découvrir des  
modèles, des tendances et  
des relations dans les  
données.

---



## Introduction aux maladies chroniques

- Introduction aux caractéristiques des maladies chroniques.
- Analyse des variables médicales présentes dans les ensembles de données.
- Etat de l'Art : Surveillance des méthodes d'analyse utilisées dans le domaine médical.



## Exploitation des données

Langage utilisé : Python et ses bibliothèques. (Pandas etc)  
Traitement des données : Jupiter Notebook

1. Exploration
2. Nettoyage

# PROCESS EDA

## Exploitation des données

### Exploration & Nettoyage

Le processus de nettoyage des données lors de l'analyse de données via Python peut être divisé en plusieurs étapes, qui sont généralement effectuées de manière séquentielle.

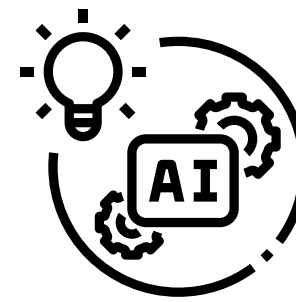


- 1 VALEURS MANQUANTES**  
Peuvent être supprimées, imputées (par exemple, remplacées par la moyenne, la médiane ou encore traitées d'une autre manière)
- 2 VALEURS ABERRANTES**  
Selon le contexte, ces valeurs peuvent être supprimées ou ajustées afin de ne pas fausser les résultats de l'analyse.
- 3 TRAITEMENT DES DOUBLONS**  
Il est important de détecter et de supprimer les observations en double dans les données puisqu'ils peuvent fausser les résultats.
- 4 CONVERSION DES TYPES DE DONNÉES**  
Il est parfois nécessaire de convertir les types de données des variables pour qu'elles soient compatibles avec les opérations d'analyse.
- 5 NORMALISATION OU MISE À L'ÉCHELLE**  
Dans certains cas, il est nécessaire de normaliser ou de mettre à l'échelle les variables afin de garantir que les différentes variables contribuent de manière égale à l'analyse.

---

## Machine Learning et création de l'application à destination des professionnels.

Intégration de modèles d'apprentissage automatique pour fournir des fonctionnalités intelligentes telles que la prédiction, la classification, l'automatisation de tâches et la recommandation, répondant ainsi aux besoins spécifiques des utilisateurs professionnels et améliorant l'efficacité opérationnelle.



## Evaluation des modèles de Machine Learning

- Random Forest (Random Forest Classifier)
- Modèle Linéaire (Régression Logistique)
- SVM (SVC - Support Vector Classifier)
- Réseau de Neurones (MLP Classifier)
- XGBoost (XGB Classifier)
- Light GBM (LGBM Classifier)
- CatBoost (CatBoost Classifier)
- Naive Bayes (Gaussian NB)

## Lancement de l'Application

Utilisation de l'application Streamlit

# PROCESS ML

## Machine Learning

### Exploration des modèles d'apprentissage

Tester et comparer différents algorithmes et configurations de modèles afin d'identifier celui qui offre les meilleures performances de prédictions.



1

#### **PARTITIONNEMENT DES DONNÉES**

Diviser l'ensemble de données en ensembles d'apprentissage, de validation et de test pour évaluer les performances du modèle.

2

#### **CHOIX DU MODÈLE AVEC PIPELINE**

Sélectionner le type de modèle adapté au problème, ici des modèles de classifications.

3

#### **OPTIMISATION DES HYPERPARAMÈTRES**

Pour améliorer ses performances, par exemple à l'aide de méthodes de recherche d'hyperparamètres comme la recherche en grille (Gridsearch).

4

#### **ENTRAÎNEMENT DU MODÈLE**

Appliquer les données d'apprentissage au modèle choisi et ajuster ses paramètres afin qu'il puisse apprendre à partir des données.

5

#### **ÉVALUATION FINALE**

Sur l'ensemble de test, qui n'a pas été utilisé pendant l'entraînement ni la validation, pour obtenir une estimation impartiale de la capacité du modèle à généraliser sur de nouvelles données.

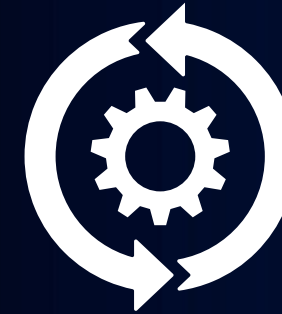


# L'APPLICATION

# EVOLUTIONS

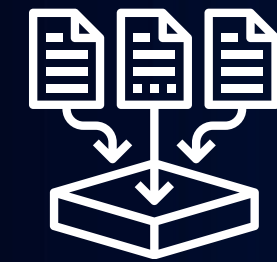
---

Évoluer une application est essentiel pour rester compétitif, répondre aux besoins des utilisateurs et garantir la satisfaction continue.



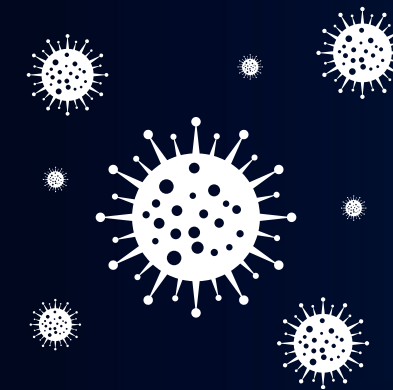
## Amélioration continue

Amélioration continue des algorithmes de machine learning pour chacune des maladies.



## Intégration des données

Connexion et synchronisation de différentes sources de données externes avec l'application.



## Nouvelles Maladies

Autres maladies chroniques, neurologiques, respiratoires cancers etc.



## Partenariats

Avec des entreprises, associations professionnelles ou experts du secteur.



DATA ANALYSIS

# MERCI

DE VOTRE ATTENTION

MEDICAL PROFESSIONAL APP

2024

