



PROJET DE FIN D'ANNEE
5ème Année en Ingénierie Informatique et Réseaux

HealthInSight

Réalisé par :

*Wissal YAGHCHA
&
Imad BENRITOUNIA*

Encadrant Pédagogique : *Abderrahim Larhlimi*

Année universitaire : 2025/2026

Remerciements

Nous tenons à exprimer nos sincères remerciements à notre encadrant académique, **Monsieur Abderrahim Larhlimi**, pour l'accompagnement rigoureux et le suivi attentif qu'il a assurés tout au long de la réalisation de ce projet de fin d'année. Grâce à son expertise scientifique, à la pertinence de ses orientations méthodologiques et à la qualité de ses conseils, nous avons pu structurer notre travail de manière cohérente et approfondir les aspects techniques et conceptuels de notre projet.

Ses remarques constructives, sa disponibilité constante et son sens pédagogique nous ont permis de mieux appréhender les enjeux liés à notre sujet, d'adopter une démarche méthodologique adaptée et d'améliorer continuellement la qualité de notre travail. Son encadrement a constitué un réel apport, tant sur le plan académique que sur le plan professionnel.

Nous souhaitons également adresser nos **vifs remerciements à Monsieur Abdelali ZAKRANI**, notre professeur de **qualité logicielle**, pour la richesse de ses enseignements et pour l'importance qu'il accorde aux bonnes pratiques de développement logiciel. À travers ses cours, ses explications claires et son exigence méthodologique, il nous a permis de mieux comprendre les enjeux de la qualité du code, de l'analyse statique et de l'utilisation d'outils professionnels tels que **SonarQube**, éléments clés dans la réalisation et l'évaluation de notre projet.

Enfin, nous lui exprimons notre profonde reconnaissance pour la confiance qu'il nous a accordée, ainsi que pour l'encouragement et le soutien continu qu'il nous a apportés tout au long de ce travail, contribuant de manière déterminante à la réussite de ce projet.

Résumé

Ce projet de fin d'année s'inscrit dans le cadre de la valorisation des données de santé publiques à travers l'intelligence artificielle et les technologies web modernes. Il vise à concevoir et développer une plateforme intelligente de prédiction du risque de maladies, basée sur l'exploitation de données ouvertes issues de Statistique Canada. Le projet met l'accent sur la prévention et l'aide à la décision en santé, en proposant une approche accessible et visuelle pour l'utilisateur final.

La solution développée permet à l'utilisateur de saisir des informations liées à son profil de santé (âge, habitudes de vie, indicateurs physiologiques). Ces données sont traitées par un modèle de machine learning capable d'estimer un niveau de risque associé à une pathologie donnée. Les résultats sont ensuite restitués via une interface web interactive intégrant une visualisation anatomique en trois dimensions, mettant en évidence les zones du corps potentiellement à risque.

Cette plateforme vise à faciliter la compréhension des données de santé, à sensibiliser les utilisateurs aux facteurs de risque et à démontrer l'apport de l'intelligence artificielle dans la prévention médicale. Elle combine des compétences en data science, machine learning, développement web et visualisation avancée.

Mots-clés : données de santé, intelligence artificielle, machine learning, prédiction médicale, visualisation 3D, prévention

Abstract

This end-of-year project is part of the exploitation of public health data through artificial intelligence and modern web technologies. Its objective is to design and develop an intelligent platform for disease risk prediction, based on open health datasets provided by Statistics Canada. The project focuses on prevention and decision support in healthcare by offering an accessible and visual approach for end users.

The developed solution allows users to enter health-related information such as age, lifestyle habits, and physiological indicators. These data are processed by a machine learning model that estimates the risk level associated with a specific disease. The prediction results are displayed through an interactive web interface featuring a three-dimensional anatomical visualization that highlights the potentially affected body areas.

This platform aims to improve the understanding of health data, raise awareness about risk factors, and demonstrate the contribution of artificial intelligence to medical prevention. It combines data science, machine learning, web development, and advanced visualization techniques.

Keywords: health data, artificial intelligence, machine learning, disease prediction, 3D visualization, prevention

Table des matières

Remerciements	2
Résumé	3
Abstract.....	4
Table des matières.....	5
Liste des figures	7
Introduction générale	8
Chapitre I : Contexte général du projet.....	10
1. Introduction	12
2. Contexte du projet.....	12
3. Problématique	13
4. Étude et critique de l'existant	13
8. Conclusion.....	16
Chapitre II : Spécification des besoins	23
& Conception.....	23
1. Introduction	24
2. Analyse.....	24
2.1. Spécifications fonctionnelles	24
2.2. Spécifications non fonctionnelles.....	25
2.3. Diagramme de contexte.....	26
3. Conception.....	27
3.1. Diagramme de cas d'utilisation	27
3.2. Diagramme de classes	27
3.3. Diagramme de séquences	28

3.4. Diagrammes d'activité.....	29
5. Conclusion.....	32
Chapitre III: Réalisation du Projet	33
1. Introduction.....	34
2. Environnement de travail	34
2.1. Environnement matériel.....	34
Webographie	45

Liste des figures

Figure 1- Les missions du projet	15
Figure 2: Besoins fonctionnelles	25
Figure 3: Besoins non fonctionnels	26
Figure 4: Diagramme de contexte	26
Figure 5: Diagramme de cas d'utilisation général	27
Figure 6: Diagramme de classes	28
Figure 7: Diagramme de séquence	29
Figure 8: Diagramme d'activité	30
Figure 9: Architecture logiciel	31
Figure 10: Architecture matérielle	32
Figure 11: Logo d'environnement de développement	35
Figure 12: Page d'accueil – HealthInsight Web	35
Figure 13: Page de connexion (Login)	36
Figure 14: Page de création de compte (Sign Up)	37
Figure 15: Formulaire de prédiction du risque de santé	37
Figure 16: Dashboard de prédiction avec visualisation 3D	38
Figure 17: Page d'exploration des données	39
Figure 18: Logo de SonarQube	39
Figure 19: Quality Gate	40
Figure 20: Analyse de sécurité et fiabilité	41
Figure 21: Analyse de maintenabilité	42
Figure 22: Anomalie détectée	42
Figure 23: Résultat final	43

Introduction générale

Ce projet de fin d'année s'inscrit dans un contexte marqué par la croissance continue des données de santé et par l'importance croissante accordée à la prévention médicale. Les organismes publics produisent aujourd'hui de vastes volumes de données issues d'enquêtes de santé, offrant une opportunité importante pour analyser les facteurs de risque et anticiper certaines maladies. Toutefois, ces données restent souvent difficiles à exploiter et peu accessibles aux utilisateurs non spécialisés.

L'objectif de ce projet est de proposer une solution innovante permettant d'exploiter ces données à travers un modèle de prédiction basé sur le machine learning, tout en offrant une restitution claire et intuitive des résultats. La plateforme développée repose sur une architecture moderne intégrant une API de prédiction et une interface web interactive.

En mettant l'accent sur la visualisation anatomique en trois dimensions, le projet vise à améliorer la compréhension des risques de santé et à sensibiliser les utilisateurs à l'importance des habitudes de vie. Cette approche permet de relier les résultats prédictifs à une représentation concrète du corps humain.

Chapitre I : Contexte général du projet

Ce chapitre présente le contexte global du projet ainsi que les limites des approches traditionnelles utilisées dans l'exploitation des données de santé. Malgré la disponibilité croissante de données de santé publiques issues d'organismes officiels, celles-ci sont souvent analysées de manière descriptive, sans intégration de modèles prédictifs ni outils de visualisation adaptés. Cette situation limite leur exploitation à des fins de prévention et d'aide à la décision.

Le chapitre met également en évidence les objectifs de la solution proposée, à savoir l'exploitation intelligente des données de santé, la prédiction du risque de maladies à l'aide du machine learning et la restitution des résultats via une interface web interactive. Il aborde enfin la démarche méthodologique adoptée, basée sur une organisation itérative du travail, ainsi que les principales contraintes techniques liées au traitement des données, à la performance des modèles et à la sécurité des échanges entre les composants du système.

Chapitre II : Analyse et conception

Ce chapitre est consacré à l'analyse des besoins fonctionnels et techniques du projet ainsi qu'à la conception de la solution. Les besoins fonctionnels portent principalement sur la saisie des données de santé, l'exécution de la prédiction du risque de maladie et la visualisation claire des résultats pour l'utilisateur. Les besoins non fonctionnels concernent notamment la performance du système, la fiabilité des prédictions et l'ergonomie de l'interface.

La conception s'appuie sur une architecture modulaire intégrant un composant de traitement des données, un modèle de machine learning, une API de prédiction et une interface web. Des schémas de conception permettent de structurer les interactions entre ces composants et de garantir la cohérence globale de la solution.

Chapitre III : Réalisation du Projet

Ce chapitre décrit la mise en œuvre technique de la plateforme développée. L'environnement de développement repose sur **VS Code** pour le backend Python et le frontend web. La solution utilise une stack technologique moderne adaptée aux projets data et IA.

La stack mise en place comprend :

- **Python** pour le traitement des données et l'entraînement du modèle de machine learning ;
- **Scikit-learn** pour la construction du modèle de prédiction du risque de maladies ;
- **FastAPI** pour l'exposition de l'API REST de prédiction ;
- **React.js** pour le développement de l'interface utilisateur ;
- **React Three Fiber** pour la visualisation anatomique en trois dimensions.

Les principales interfaces développées incluent un formulaire de saisie des données de santé et une vue de visualisation 3D permettant de représenter les zones du corps associées au risque prédictif. Des tests fonctionnels ont été réalisés afin de valider le bon fonctionnement des principaux cas d'usage, depuis la saisie des données jusqu'à l'affichage des résultats.

Chapitre IV : Conclusion et Perspectives

Ce chapitre présente un bilan du projet réalisé et des objectifs atteints. La plateforme développée permet d'exploiter des données de santé publiques, de prédire le risque de maladies à l'aide de modèles de machine learning et de restituer les résultats de manière visuelle et intuitive. Le projet démontre l'intérêt de combiner intelligence artificielle, data engineering et visualisation avancée dans le domaine de la santé.

Plusieurs perspectives d'évolution peuvent être envisagées, notamment :

- l'amélioration des modèles de prédiction par l'intégration de nouvelles variables de santé ;
- l'ajout de modules explicatifs pour rendre les prédictions plus interprétables ;
- l'intégration de tableaux de bord statistiques pour le suivi global des risques ;
- le déploiement de la solution dans un environnement cloud afin d'améliorer sa scalabilité.

Ce projet constitue ainsi une première étape vers des solutions intelligentes orientées prévention et aide à la décision en santé, avec une valeur ajoutée tant académique que pratique.

Chapitre I : Contexte général du projet

Chapitre I : Contexte général du projet

1. Introduction

Ce chapitre présente le contexte fonctionnel et technique du projet, ainsi que les objectifs poursuivis dans le cadre du développement de la plateforme de prédiction en santé. Bien que les données de santé publiques soient largement disponibles, leur exploitation reste souvent limitée à des analyses statistiques classiques, sans outils interactifs ni prédictifs accessibles au grand public.

Le projet vise ainsi à transformer ces données brutes en informations exploitables, en combinant traitement des données, intelligence artificielle et visualisation interactive. Il s'inscrit dans une démarche de modernisation de l'exploitation des données de santé.

2. Contexte du projet

Les données de santé issues des enquêtes nationales, telles que celles de Statistique Canada, couvrent un large éventail d'informations relatives aux habitudes de vie, aux facteurs de risque et aux maladies chroniques. Cependant, ces données sont souvent complexes, volumineuses et difficiles à interpréter sans outils adaptés.

La plateforme développée dans ce projet a pour objectif de :

- permettre l'exploitation simplifiée des données de santé publiques ;
- offrir un modèle de prédiction du risque de maladies basé sur le machine learning ;
- fournir une interface web intuitive pour visualiser les résultats ;
- améliorer l'accessibilité et la compréhension des informations de santé.

L'intégration d'une visualisation 3D constitue un élément différentiateur du projet, permettant de relier les prédictions à des zones anatomiques précises.

3. Problématique

Malgré la disponibilité de données de santé ouvertes et fiables, leur exploitation reste limitée par l'absence de solutions intégrées combinant prédition et visualisation. Les analyses existantes sont souvent réservées aux experts et peu adaptées à une utilisation pédagogique ou préventive.

La problématique centrale du projet peut donc être formulée comme suit : **Comment exploiter efficacement des données de santé ouvertes afin de prédire les risques de maladies et de restituer les résultats de manière intuitive, visuelle et compréhensible pour l'utilisateur final ?**

Cette problématique soulève des enjeux techniques (traitement des données, modélisation, performance), mais également des enjeux fonctionnels liés à l'expérience utilisateur et à l'interprétation des résultats.

4. Étude et critique de l'existant

L'étude de l'existant met en évidence plusieurs insuffisances dans la gestion et l'exploitation des données de santé à des fins d'analyse et de prédition. Les solutions actuellement disponibles reposent majoritairement sur des outils statistiques classiques ou sur des tableaux de bord descriptifs, offrant une vision globale mais peu personnalisée de l'état de santé des populations. Ces approches se limitent souvent à la consultation de données brutes ou d'indicateurs agrégés, sans réel mécanisme d'analyse prédictive ou d'aide à la décision.

De plus, l'accès aux données de santé publiques, bien que facilité par des organismes officiels tels que Statistique Canada, reste complexe pour les utilisateurs non techniques. Les données sont généralement diffusées sous forme de fichiers volumineux (CSV, ZIP), nécessitant des compétences en traitement de données pour être exploitées efficacement. L'absence de processus ETL automatisé et structuré rend leur utilisation fastidieuse et peu adaptée à un usage interactif.

Enfin, les solutions existantes intègrent rarement des mécanismes de visualisation avancée permettant une interprétation intuitive des résultats. L'utilisateur est souvent confronté à des graphiques statiques ou à des tableaux numériques, sans représentation visuelle directe de l'impact des prédictions sur le corps humain. Ces limites réduisent la valeur ajoutée des outils existants, tant sur le plan pédagogique que décisionnel.

5. Solution proposée

Face aux limites identifiées, le projet *Healthinsight* propose une solution innovante reposant sur la conception d'une application web intelligente dédiée à l'analyse et à la prédition d'indicateurs de santé. Contrairement aux solutions existantes, cette application intègre un pipeline complet allant de l'extraction automatisée des données de santé publiques jusqu'à la restitution des résultats sous une forme visuelle et interactive.

La solution repose sur l'automatisation du processus ETL afin de transformer des jeux de données bruts issus de Statistique Canada en données exploitables par des modèles de machine learning. Ces modèles permettent de réaliser des prédictions ou des estimations de risques à partir de variables de santé pertinentes. Les résultats sont ensuite exposés via une API backend et présentés à l'utilisateur à travers une interface web ergonomique.

L'originalité du projet réside dans l'intégration d'une visualisation 3D du corps humain, permettant de représenter de manière intuitive les zones ou systèmes potentiellement concernés par les prédictions. Cette approche améliore considérablement la compréhension des résultats par l'utilisateur final.

Parmi les avantages de la solution proposée figurent l'automatisation, l'interactivité et la valeur pédagogique. En revanche, elle dépend de la qualité et de la granularité des données publiques disponibles, ce qui constitue une limite partielle. Malgré cela, le choix de cette solution est justifié par son caractère innovant, sa pertinence académique et son applicabilité dans un contexte réel.

6. Choix du modèle de développement

Pour la réalisation du projet *Healthinsight*, une démarche de développement itérative et incrémentale a été adoptée, inspirée du modèle Agile. Ce choix se justifie par la nature exploratoire du projet, combinant traitement de données, intelligence artificielle et développement web, où les besoins peuvent évoluer au fur et à mesure de l'avancement.

Le modèle Agile permet de découper le projet en phases successives, chacune aboutissant à un résultat exploitable (pipeline de données, modèle de prédition, API, interface utilisateur). Cette approche favorise la validation progressive des fonctionnalités, la détection précoce des problèmes techniques et l'adaptation rapide aux contraintes temporelles du cadre académique.

Comparé aux modèles classiques en cascade, le modèle Agile offre une meilleure flexibilité et une meilleure maîtrise des risques, notamment dans un projet intégrant des technologies hétérogènes. Il constitue ainsi un choix pertinent pour assurer la cohérence, la qualité et la réussite du projet.

7. Planning prévisionnel

L'image ci-dessous illustre les différentes missions du projet.

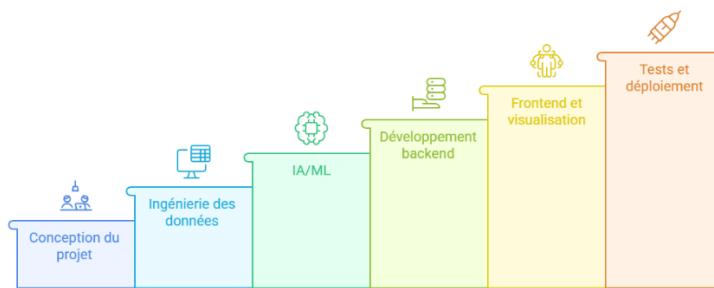


Figure 1- Les missions du projet

Cette démarche globale couvre l'ensemble du cycle de développement du projet *Healthinsight*, depuis la phase de conception jusqu'au déploiement et à la documentation finale. Elle débute par l'identification de la problématique dans le domaine de la santé et la définition des objectifs du projet, suivies par la sélection de jeux de données de santé issus de sources officielles telles que Statistique Canada et la conception de l'architecture globale du système.

La phase d'ingénierie des données repose sur la mise en place d'un processus ETL structuré, incluant l'extraction des données, leur nettoyage, leur transformation et leur préparation pour l'analyse. Ces données servent ensuite de base à la phase d'intelligence artificielle, au cours de laquelle des modèles de machine learning sont sélectionnés, entraînés et évalués afin de fournir des prédictions fiables liées aux indicateurs de santé.

Le backend du système intègre ces modèles au sein d'une API développée avec FastAPI, assurant la gestion des requêtes de prédiction, la validation des données et la sécurisation des échanges. En parallèle, une interface web ergonomique est développée afin de permettre la saisie des données utilisateur, l'affichage des résultats de prédiction et leur représentation visuelle à travers une modélisation 3D du corps humain mettant en évidence les zones concernées.

Enfin, des phases de tests fonctionnels et de validation du modèle garantissent la fiabilité de la solution, avant un déploiement local accompagné d'une documentation complète et de la rédaction du rapport final. Cette approche structurée assure la cohérence entre les aspects métiers, techniques et scientifiques du projet, tout en répondant aux exigences académiques d'un projet de fin d'études.

8. Conclusion

Ce chapitre a permis de présenter une analyse critique de l'existant, mettant en évidence les limites des solutions actuelles en matière d'exploitation et de prédition des données de santé. À partir de cette analyse, une solution innovante et adaptée a été proposée, reposant sur l'automatisation du traitement des données, l'utilisation de modèles de machine learning et une restitution visuelle avancée des résultats.

Le choix du modèle de développement Agile ainsi que l'élaboration d'un planning prévisionnel structuré constituent des éléments clés pour assurer la réussite du projet *Healthinsight*. Les objectifs définis dans ce chapitre servent de fondement aux phases suivantes, dédiées à la conception détaillée et à la réalisation technique de la solution proposée.

*Chapitre II : Spécification des besoins
& Conception*

1. Introduction

La spécification des besoins constitue une étape essentielle dans la conception de toute application informatique, car elle permet de définir clairement les attentes fonctionnelles et techniques du système à développer. Dans le cadre du projet *Healthinsight*, cette phase vise à identifier les besoins des utilisateurs ainsi que les exigences que la plateforme doit satisfaire afin d'assurer une exploitation efficace des données de santé et une restitution pertinente des résultats de prédiction.

Ce chapitre a pour objectif de formaliser les besoins fonctionnels et non fonctionnels de la solution proposée. Il sert de référence pour les phases de conception et de réalisation, en garantissant la cohérence entre les objectifs initiaux du projet et les fonctionnalités effectivement développées.

2. Analyse

2.1. Spécifications fonctionnelles

Les besoins fonctionnels décrivent l'ensemble des services et fonctionnalités que la plateforme *Healthinsight* doit offrir à ses utilisateurs. Ces besoins sont directement liés aux objectifs du projet et aux usages attendus de l'application.

La plateforme doit permettre à l'utilisateur de saisir des informations relatives à son profil de santé, telles que l'âge, le sexe, les habitudes de vie et certains indicateurs physiologiques. À partir de ces données, le système doit être capable de lancer un processus de prédiction reposant sur un modèle de machine learning afin d'estimer un niveau de risque associé à une pathologie donnée.

Le système doit également assurer l'affichage clair et compréhensible des résultats de prédiction. Ces résultats incluent le niveau de risque estimé ainsi que leur représentation visuelle à travers une modélisation en trois dimensions du corps humain, mettant en évidence les zones concernées. Enfin, la plateforme doit offrir une navigation fluide et intuitive, permettant à l'utilisateur d'interagir facilement avec les différentes fonctionnalités proposées.

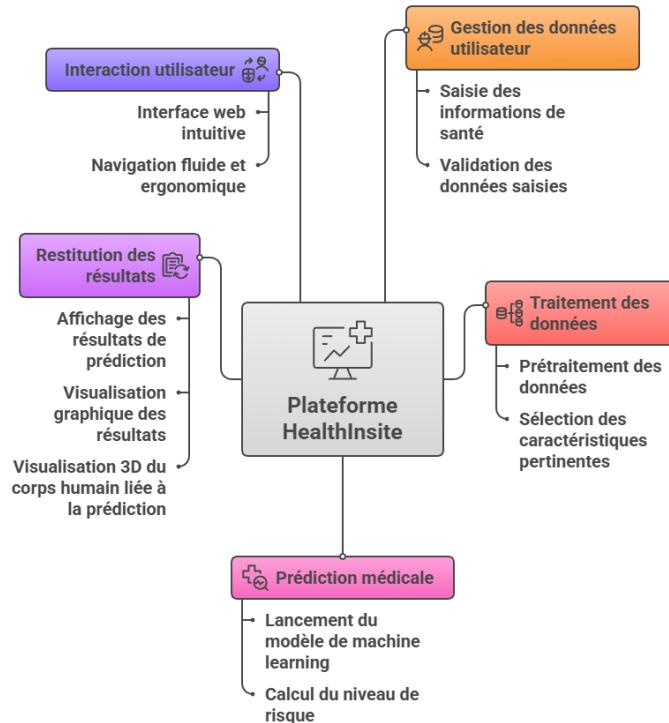


Figure 2: Besoins fonctionnelles

2.2. Spécifications non fonctionnelles

Les besoins non fonctionnels définissent les contraintes et les critères de qualité auxquels la plateforme *Healthinsight* doit répondre. Ils concernent principalement les aspects liés à la performance, à la sécurité, à l'ergonomie et à la fiabilité du système.

En termes de performance, l'application doit être capable de fournir une réponse rapide aux requêtes de prédition afin d'assurer une expérience utilisateur satisfaisante. La fiabilité du système est également un critère essentiel, impliquant une gestion appropriée des erreurs et une validation des données saisies par l'utilisateur.

La sécurité constitue un autre besoin non fonctionnel majeur, notamment en ce qui concerne la protection des données de santé saisies et la sécurisation des échanges entre le frontend et le backend.

Enfin, l'ergonomie de l'interface doit être pensée de manière à rendre l'application accessible à des utilisateurs non experts, grâce à une interface claire, intuitive et cohérente.

Chapitre II : Spécification des besoins & conception

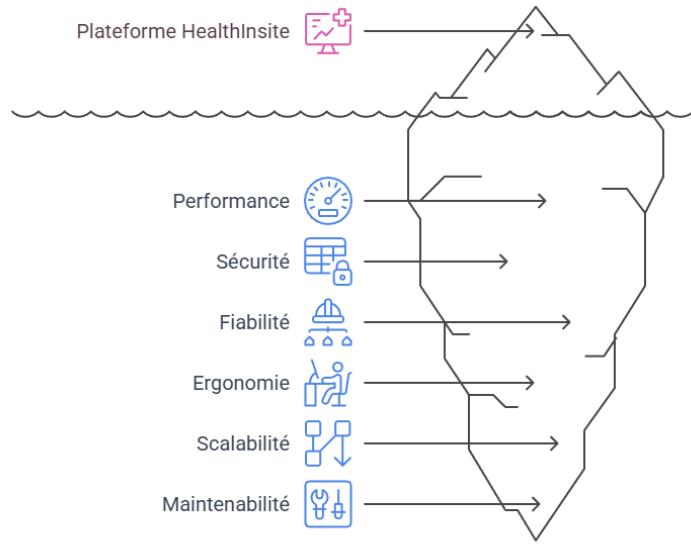


Figure 3: Besoins non fonctionnels

2.3. Diagramme de contexte

Avant de détailler les fonctionnalités internes du système, il est essentiel de situer le projet dans son environnement global. Le diagramme de contexte permet de représenter, à un niveau macro, les principaux acteurs qui interagissent avec la plateforme HealthInsight ainsi que les systèmes externes mobilisés (sources de données, services backend, modèle de prédiction, base de données et module de visualisation). Il offre une vue synthétique des échanges d'informations et clarifie les frontières du système, ce qui facilite la compréhension de l'architecture générale avant d'aborder les diagrammes plus détaillés.

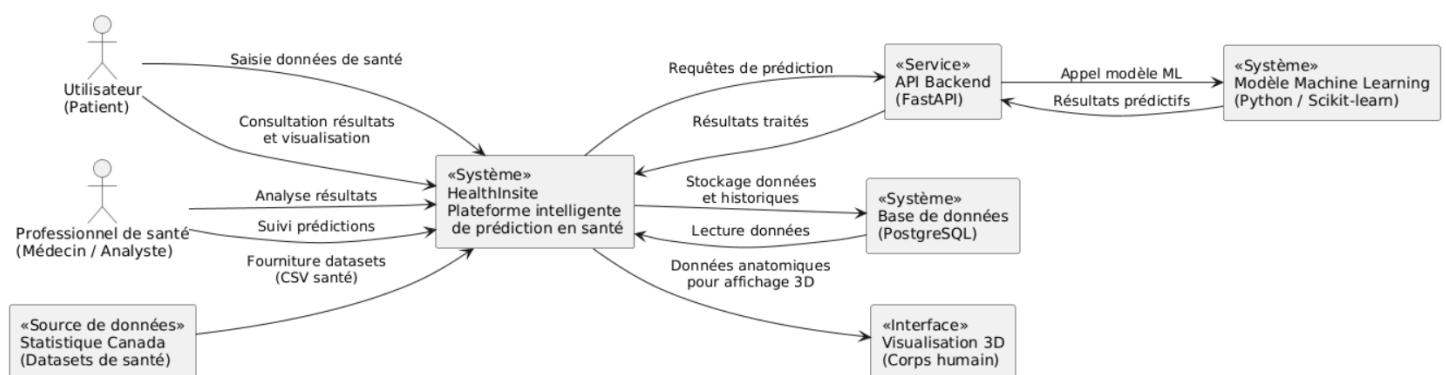


Figure 4: Diagramme de contexte

3. Conception

3.1. Diagramme de cas d'utilisation

Le diagramme de cas d'utilisation décrit les principales fonctionnalités offertes par Healthinsite du point de vue des utilisateurs. Il met en évidence les interactions entre les acteurs (patient/utilisateur, professionnel de santé/analyste) et le système, en présentant les services accessibles : saisie des informations, lancement de prédiction, consultation des résultats, et visualisation 3D. Ce diagramme sert de base à la spécification fonctionnelle en identifiant clairement les besoins et les responsabilités de chaque acteur.

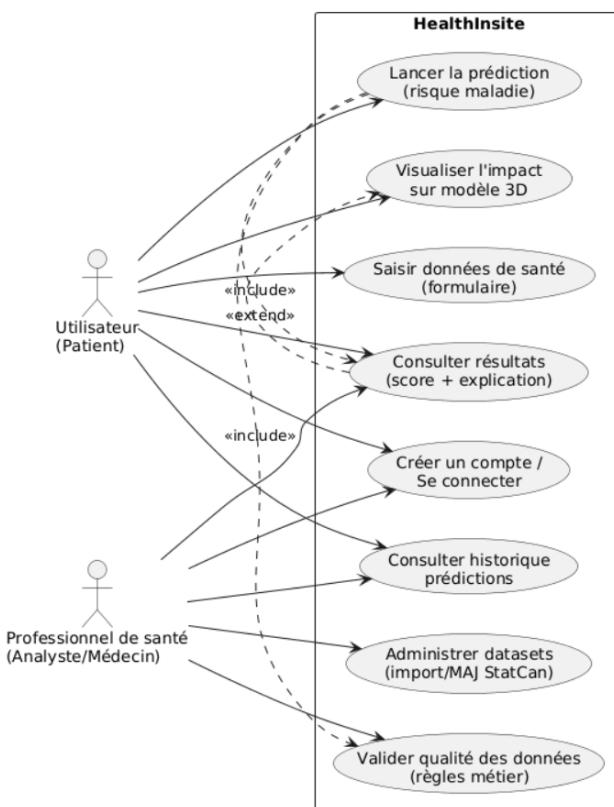


Figure 5: Diagramme de cas d'utilisation général

3.2. Diagramme de classes

Le diagramme de classes présente la structure statique du système Healthinsight. Il décrit les principales entités manipulées par la plateforme : utilisateur, enregistrement de données de santé, résultat de prédiction et dataset importé. Ce diagramme clarifie les attributs essentiels,

Chapitre II : Spécification des besoins & conception

les relations entre classes (composition, association) et constitue une base utile pour l'implémentation côté backend ainsi que la conception de la base de données.

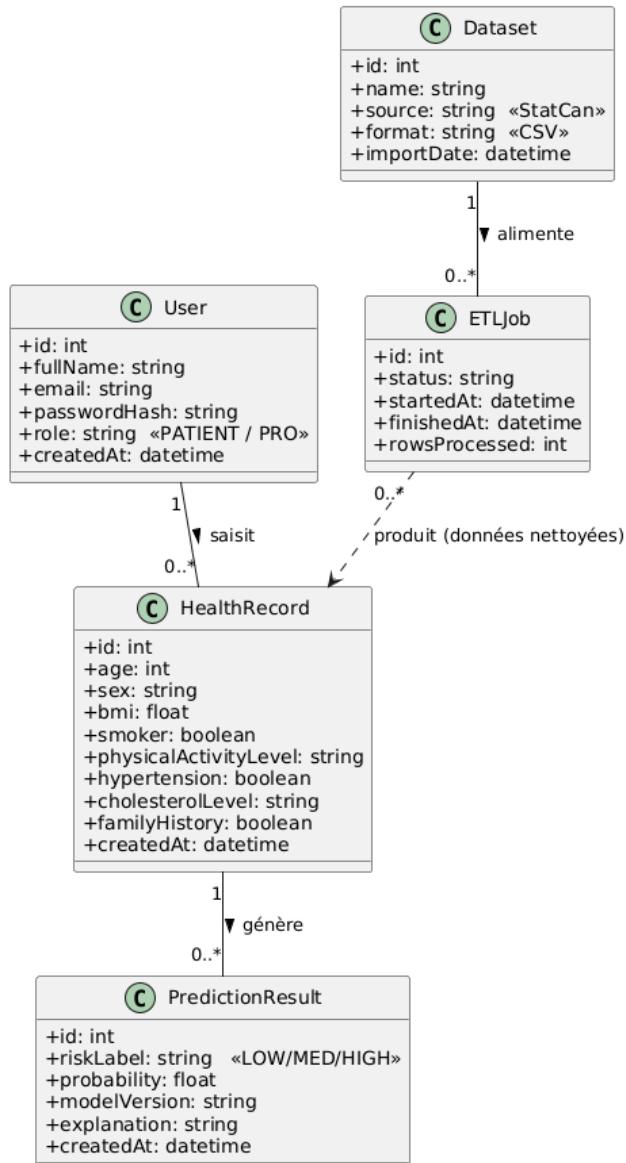


Figure 6: Diagramme de classes

3.3. Diagramme de séquences

Le diagramme de séquence illustre le déroulement temporel d'un scénario clé : la prédiction d'un risque de maladie à partir de données saisies par l'utilisateur. Il montre comment l'interface web transmet la requête au backend (FastAPI), comment les données sont validées et prétraitées, puis comment le modèle de machine learning est invoqué pour produire une prédiction. Le résultat est ensuite enregistré dans la base de données et renvoyé à l'utilisateur

Chapitre II : Spécification des besoins & conception

avec une visualisation 3D associée.

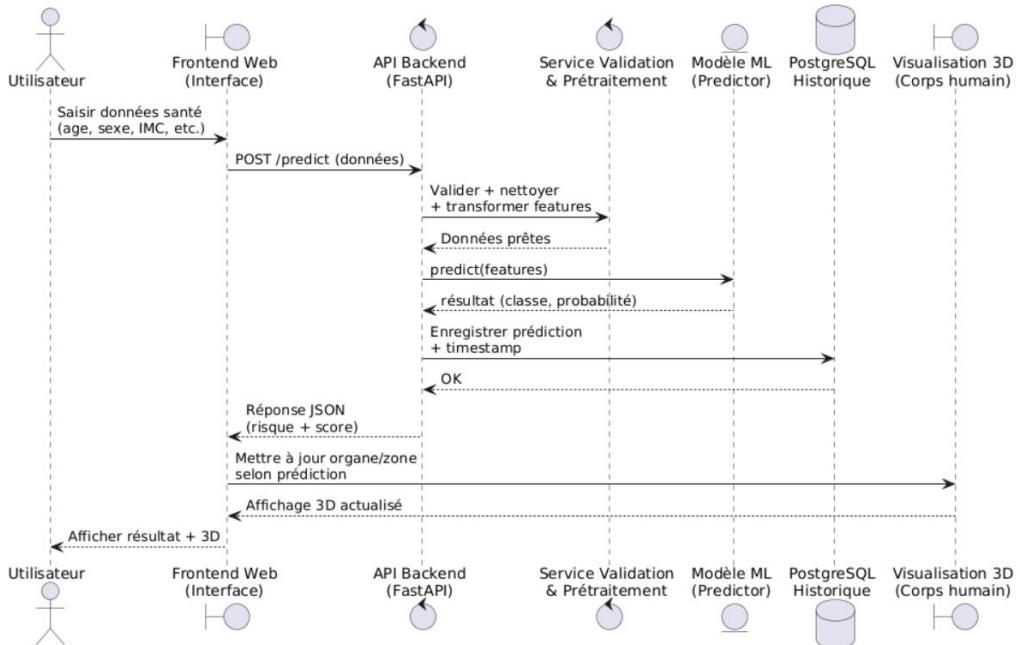


Figure 7: Diagramme de séquence

3.4. Diagrammes d'activité

Le diagramme d'activité ci-dessous décrit le déroulement global du processus de prédiction au sein de la plateforme *Healthinsight*. Il met en évidence les différentes étapes depuis l'interaction initiale de l'utilisateur avec l'application, en passant par la saisie et la validation des données de santé, jusqu'à l'exécution du modèle de machine learning et l'affichage des résultats. Ce diagramme permet de visualiser de manière séquentielle et conditionnelle les traitements effectués par le système, ainsi que les décisions prises en cas de données invalides, assurant ainsi une compréhension claire du fonctionnement opérationnel de la solution.

Chapitre II : Spécification des besoins & conception

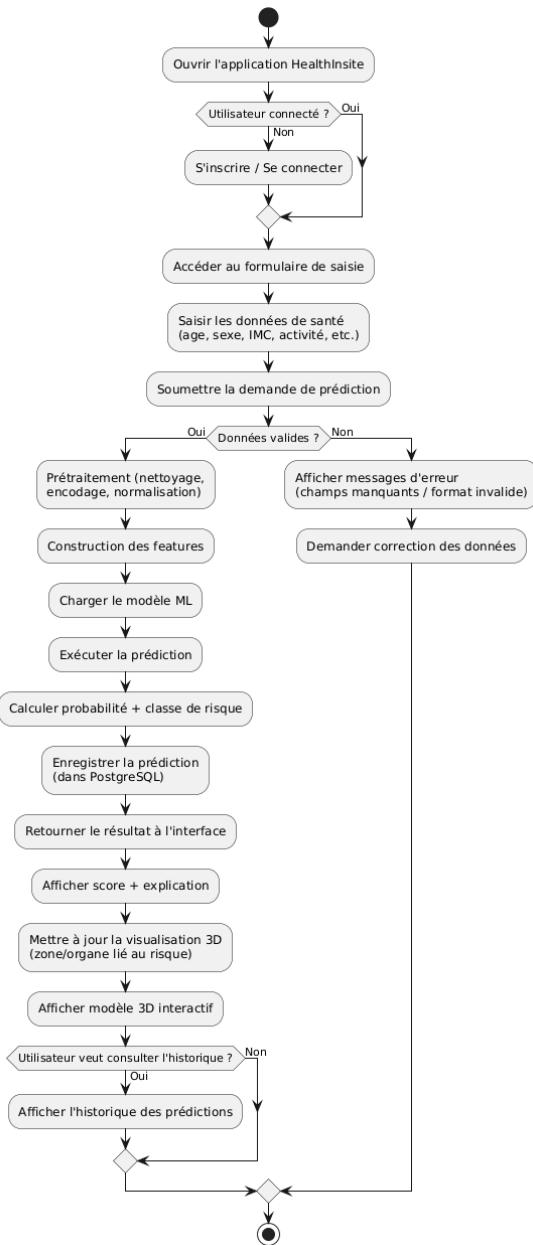


Figure 8: Diagramme d'activité

4. Architecture de l'application

4.1. Architecture logiciel

L'architecture logicielle de *Healthinsight* repose sur une approche modulaire et orientée services. Elle est organisée autour de plusieurs composants principaux, chacun ayant une responsabilité bien définie.

Le frontend web permet aux utilisateurs de saisir leurs informations de santé et de consulter les résultats de prédiction ainsi que la visualisation 3D associée. Le backend, développé avec FastAPI, assure la gestion des requêtes, la validation des données, l'orchestration des

Chapitre II : Spécification des besoins & conception

traitements et l'exposition des services de prédiction. Le module de machine learning est intégré au backend et se charge de l'entraînement, de l'évaluation et de l'exécution des modèles prédictifs à partir des données de santé. Enfin, une couche de stockage permet la gestion des jeux de données, des résultats et des métadonnées associées au projet.

Cette architecture favorise une séparation claire des responsabilités, facilite les évolutions futures (ajout de nouveaux modèles ou interfaces) et garantit une meilleure maintenabilité du système.

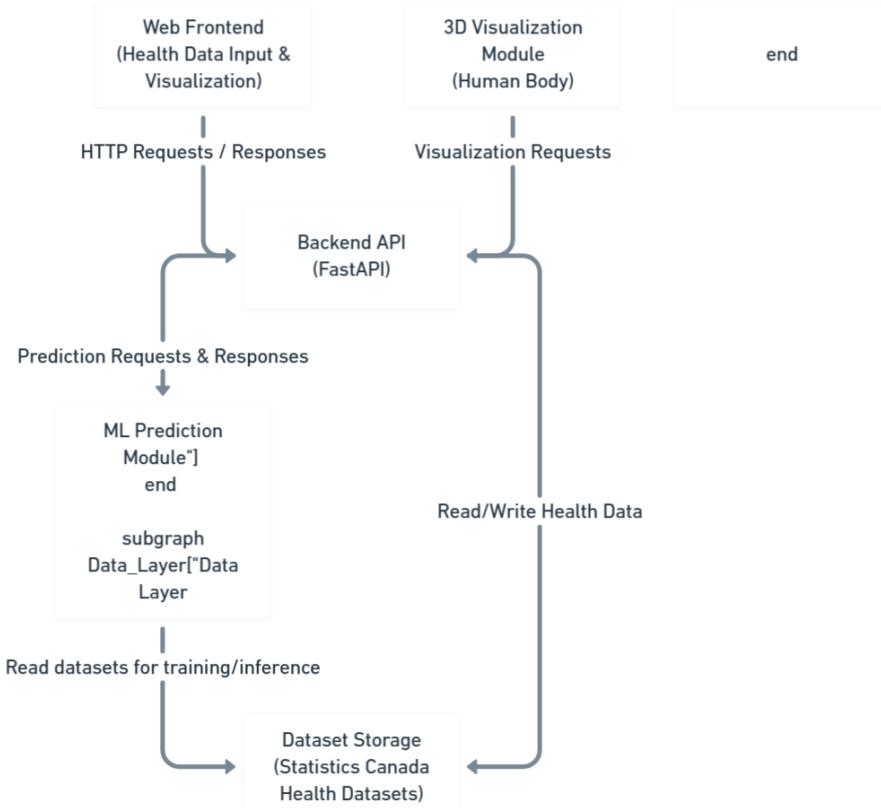


Figure 9: Architecture logiciel

4.2. Architecture matérielle

L'architecture matérielle de *Healthinsight* décrit la manière dont les composants logiciels sont déployés sur les ressources matérielles et logicielles. Dans le cadre de ce projet académique, l'application est déployée localement ou sur un serveur de développement. Le navigateur web de l'utilisateur permet d'accéder à l'interface frontend. Celui-ci communique avec un serveur applicatif hébergeant l'API FastAPI et le module de machine learning. Les données de santé, ainsi que les résultats de prédiction, sont stockés dans un espace de

Chapitre II : Spécification des besoins & conception

stockage local ou une base de données dédiée. Les échanges entre les différents nœuds s'effectuent via des protocoles de communication standards (HTTP/HTTPS).

Cette architecture matérielle assure une exécution fluide de l'application tout en restant adaptée aux contraintes académiques et aux ressources disponibles.



Figure 10: Architecture matérielle

5. Conclusion

Ce chapitre a permis de poser les fondations du projet *Healthinsight* en traduisant les objectifs fonctionnels et techniques en spécifications claires et structurées. L'identification des besoins fonctionnels et non fonctionnels, appuyée par les différents diagrammes UML, a contribué à une compréhension approfondie du fonctionnement attendu du système ainsi qu'à la définition de ses interactions avec les acteurs et les composants externes. Cette phase d'analyse et de spécification constitue un socle essentiel pour la suite du projet, en assurant une conception cohérente, sécurisée et évolutive, tout en garantissant une exploitation fiable des données de santé et une restitution pertinente des résultats de prédiction au sein de la plateforme *Healthinsight*.

Chapitre III: Réalisation du Projet

Chapitre III: Réalisation du Projet

1. Introduction

Ce chapitre est consacré à la phase de réalisation du projet **Healthinsight**. Il présente les moyens matériels et logiciels mis en œuvre afin d'aboutir à une solution fonctionnelle répondant aux objectifs définis lors des phases d'analyse et de conception. Il décrit également la mise en œuvre technique de la solution proposée ainsi que les principales interfaces graphiques développées.

Ce chapitre est structuré en deux parties principales. La première partie détaille l'environnement de développement matériel et logiciel adopté. La seconde partie est dédiée à la présentation des interfaces graphiques majeures de l'application, accompagnées d'une brève description de leur rôle et de leur contenu.

2. Environnement de travail

2.1. Environnement matériel

Le développement du projet **Healthinsight** a été réalisé sur un ordinateur personnel disposant d'une configuration matérielle suffisante pour supporter les tâches de traitement de données, d'entraînement de modèles de machine learning et de développement web.

L'environnement matériel utilisé se caractérise principalement par :

- Un processeur multi-cœurs assurant une exécution fluide des calculs et des scripts de traitement des données.
- Une mémoire vive (RAM) permettant le chargement et la manipulation des jeux de données de santé.
- Un espace de stockage suffisant pour héberger les datasets CSV, les modèles entraînés et les ressources du projet.

Le projet étant développé et testé en local, aucune infrastructure réseau spécifique (serveurs distants ou équipements réseau avancés) n'a été requise à ce stade. Toutefois, l'architecture adoptée permet une évolution future vers un déploiement sur serveur.

2.2. Environnement logiciel

L'environnement logiciel constitue un élément clé dans la réalisation du projet. Plusieurs outils et technologies ont été utilisés afin de couvrir l'ensemble du cycle de développement, depuis l'ingénierie des données jusqu'à la visualisation des résultats.

Les principaux outils et technologies utilisés sont :

- **Python** pour le traitement des données, l'implémentation du pipeline ETL et le développement des modèles de machine learning.

Chapitre III: Réalisation du Projet

- **FastAPI** pour la création d'API REST assurant la communication entre le backend et le frontend.
- **Scikit-learn** pour l'entraînement et l'évaluation des modèles de prédition.
- **Pandas et NumPy** pour la manipulation et l'analyse des données.
- **VS Code** comme environnement de développement intégré.
- **Bibliothèques de visualisation 3D** pour l'affichage interactif des prédictions sur un modèle de corps humain.

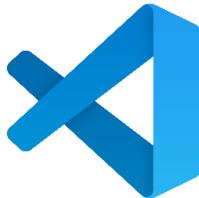


Figure 11: Logo d'environnement de développement

3. Principales interfaces graphiques

3.1. Interface d'accueil

Cette interface constitue le point d'entrée principal de la plateforme *HealthInsight Web*. Elle présente l'objectif global de l'application, à savoir la prédition des risques de santé à partir de données issues de Statistique Canada, tout en indiquant l'état de disponibilité de l'API et du modèle de machine learning. Elle permet également à l'utilisateur de naviguer facilement vers les différentes fonctionnalités clés du système.

The screenshot shows the main landing page of the HealthInsight Web application. At the top, there is a navigation bar with the title "HealthInsight Web" and links for "Accueil", "Prédiction", "Dashboard 3D", and "Exploration". The main content area features a central box titled "Bienvenue sur HealthInsight Web" with the subtitle "Plateforme web de prédition des risques de santé basée sur les données StatCan". Below this, a green button indicates "API opérationnelle - Modèle chargé". To the left is a box for "Prédiction de Risque" with a magnifying glass icon, a brief description about analyzing health profiles, and a link "Commencer l'analyse". To the right is a box for "Exploration des Données" with a bar chart icon, a brief description about visualizing statistics and trends, and a link "Explorer les données".

Figure 12: Page d'accueil – HealthInsight Web

3.2. Page de connexion

Cette interface constitue le point d'entrée principal de la plateforme *HealthInsight Web*. Elle présente l'objectif global de l'application, à savoir la prédition des risques de santé à partir de données issues de Statistique Canada, tout en indiquant l'état de disponibilité de l'API et du modèle de machine learning. Elle permet également à l'utilisateur de naviguer facilement vers les différentes fonctionnalités clés du système.

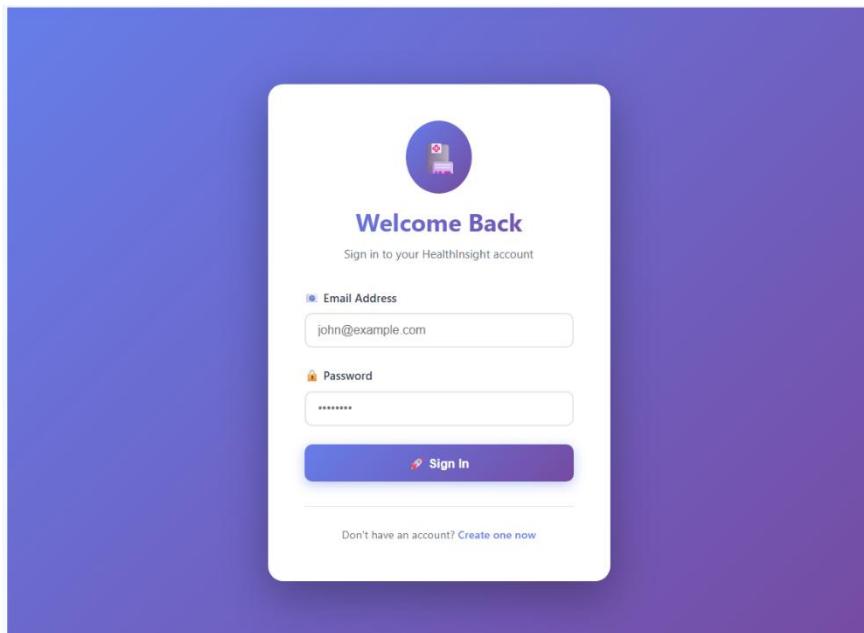


Figure 13: Page de connexion (Login)

3.3. Page de création de compte

Cette interface permet à un nouvel utilisateur de créer un compte sur la plateforme *HealthInsight*. Elle collecte les informations de base nécessaires à l'inscription et applique des règles de validation sur les champs saisis, notamment pour le mot de passe, afin d'assurer un niveau minimal de sécurité.

Chapitre VI : Conclusion et Perspectives

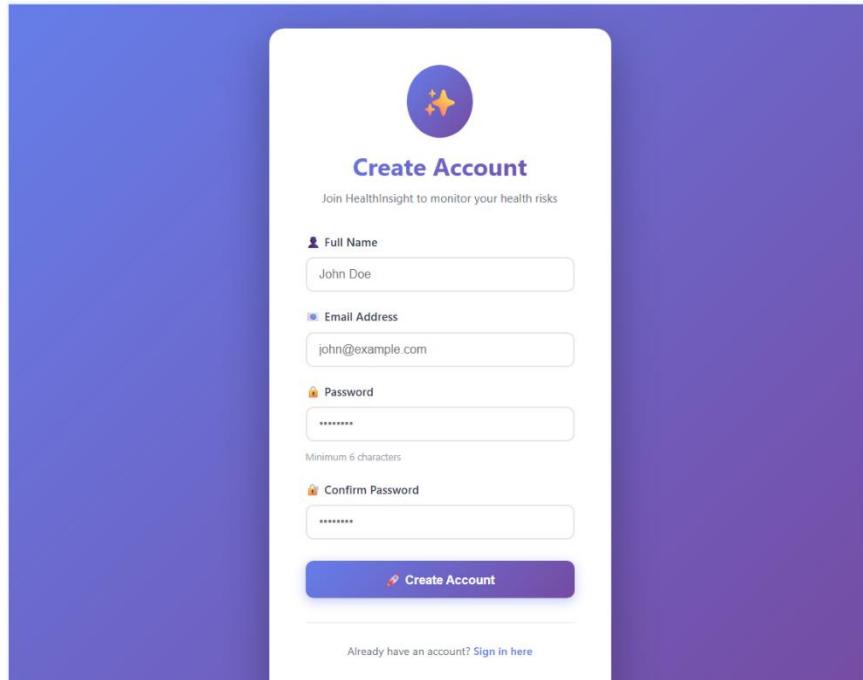


Figure 14: Page de création de compte (Sign Up)

3.4. Formulaire de prédition du risque de santé

Cette interface propose un formulaire interactif permettant à l'utilisateur de saisir ses informations de santé telles que l'âge, le sexe, l'IMC, le statut tabagique, le niveau d'activité physique et les antécédents médicaux. Ces données constituent les variables d'entrée du modèle de machine learning utilisé pour estimer le niveau de risque de santé.

A screenshot of the 'Analyse de Risque de Santé' (Risk Analysis) form. The top navigation bar includes 'HealthInsight Web', 'Accueil', 'Prédiction', 'Dashboard 3D', and 'Exploration'. The main section has a title 'Analyse de Risque de Santé' with a magnifying glass icon. A sub-instruction says 'Remplissez le formulaire ci-dessous avec vos informations de santé pour obtenir une estimation de votre niveau de risque.' Below is a 'Informations du Patient' section with various dropdown and input fields: 'Âge (ans)' set to 45, 'Sexe' set to 'Masculin', 'IMC (Indice de Masse Corporelle)' set to 25; 'Fumeur' set to 'Non', 'Niveau d'activité physique' set to 'Modéré', 'Hypertension' set to 'Non'; 'Niveau de cholestérol' set to 'Normal', 'Antécédents familiaux' set to 'Non'. A large blue 'ANALYSER LE RISQUE' button is at the bottom.

Figure 15: Formulaire de prédition du risque de santé

3.5. Dashboard de prédition avec visualisation 3D

Le dashboard de prédition affiche les résultats de l'analyse sous forme visuelle et intuitive. Une visualisation 3D du corps humain est utilisée pour représenter les zones potentiellement à risque, offrant ainsi une lecture claire et pédagogique des résultats de prédition générés par le modèle.

The screenshot shows the 'HealthInsight Web' dashboard. At the top, there's a blue header bar with the logo and navigation links: Accueil, Prédition, Dashboard 3D, and Exploration. Below the header, the main title is 'Dashboard de Prédition de Risque' with a subtitle 'Visualisation 3D des zones à risque basée sur vos données de santé'. On the left, there's a form titled 'Données Patient' containing fields for Age (45), Sexe (Masculin), IMC (25), Fumeur (Non), Activité Physique (Moyenne), and checkboxes for Hypertension, Cholestérol (Normal), and Antécédents Familiaux. A blue button at the bottom right of the form says 'Prédire le Risque'. To the right of the form is a large white area with a small 3D human body icon and the text 'En attente de prédition' followed by a sub-instruction: 'Remplissez le formulaire et cliquez sur "Prédire le Risque" pour visualiser les résultats en 3D'.

Figure 16: Dashboard de prédition avec visualisation 3D

3.6. Page d'exploration des données

Cette interface est dédiée à l'exploration statistique du dataset de santé utilisé pour l'entraînement du modèle. Elle présente des indicateurs globaux (taille de l'échantillon, taux de maladie, âge moyen, IMC moyen) ainsi que des graphiques illustrant les distributions par sexe, par état de santé et par facteurs de risque, facilitant ainsi la compréhension des données.

Chapitre VI : Conclusion et Perspectives



Figure 17: Page d'exploration des données

4. Qualité et Tests

4.1. Présentation de SonarQube

SonarQube est un outil d'analyse statique de la qualité du code source. Il permet d'évaluer automatiquement un projet logiciel selon plusieurs critères essentiels tels que la **fiabilité**, la **sécurité**, la **maintenabilité**, la **duplication du code** et la **couverture des tests**. Contrairement aux tests dynamiques, SonarQube analyse le code **sans l'exécuter**, en s'appuyant sur un ensemble de règles reconnues issues des bonnes pratiques de génie logiciel.

Dans le cadre du projet **SmartPark**, SonarQube a été utilisé afin de vérifier la conformité du code aux standards de qualité, d'identifier d'éventuelles anomalies et de garantir un niveau de qualité acceptable avant la livraison du projet.



Figure 18: Logo de SonarQube

4.2. Mise en place de l'analyse SonarQube dans le projet SmartPark

L'analyse de la qualité logicielle a été réalisée localement à l'aide de SonarQube, en intégrant l'outil au cycle de build Maven du projet. Après la configuration du projet dans SonarQube et la génération d'un token d'authentification, l'analyse a été lancée à l'aide de la commande suivante :

```
mvn clean verify sonar:sonar -Dsonar.projectKey=SmartPark -  
Dsonar.projectName=Healthinsight -Dsonar.host.url=http://localhost:9000 -  
Dsonar.token=sq_778d9ae7b6ec30cfdfc07c5d34c56a2c78fa8a23
```

Cette commande permet d'exécuter les phases de compilation, de tests, puis d'envoyer les résultats de l'analyse vers le serveur SonarQube pour une visualisation détaillée via l'interface web.

4.3. Résultats globaux de l'analyse (Quality Gate)

La première analyse du projet Healthinsight montre que le **Quality Gate est validé (Passed)**. Cela signifie que le projet respecte les seuils de qualité définis, notamment en matière de sécurité, de fiabilité et de maintenabilité.

Malgré la présence d'un avertissement mineur, aucun problème critique n'a été détecté, ce qui confirme la robustesse globale du code source.

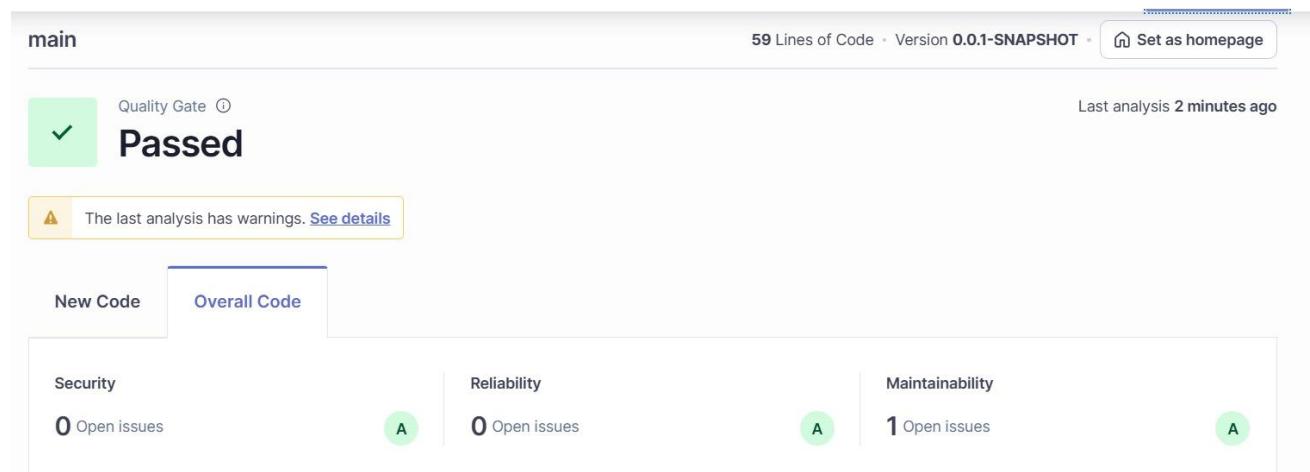


Figure 19: Quality Gate

4.4. Analyse de la sécurité et de la fiabilité

Les résultats indiquent :

- **0 vulnérabilité détectée** (Security : A)
- **0 bug détecté** (Reliability : A)

Ces résultats démontrent que le code du projet Healthinsight ne présente pas de failles de sécurité connues ni de comportements susceptibles de provoquer des erreurs d'exécution. L'outil SonarQube n'a signalé aucun problème pouvant compromettre la stabilité ou la sûreté du système.

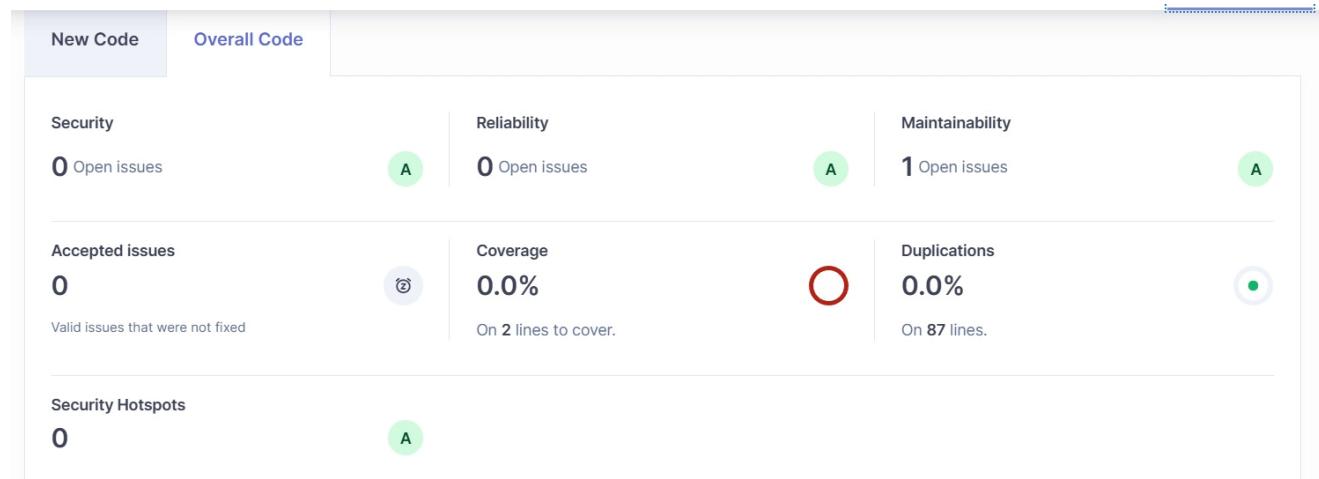


Figure 20: Analyse de sécurité et fiabilité

4.5. Analyse de la maintenabilité

Lors de la première analyse, SonarQube a détecté **une seule anomalie de type “Code Smell”**, liée à un problème de **maintenabilité**. Cette anomalie concernait la présence de **code commenté inutile** dans le fichier de configuration pom.xml.

Bien que ce type de problème ne provoque pas d'erreur fonctionnelle, il nuit à la lisibilité du code et à sa maintenance à long terme.

Chapitre VI : Conclusion et Perspectives

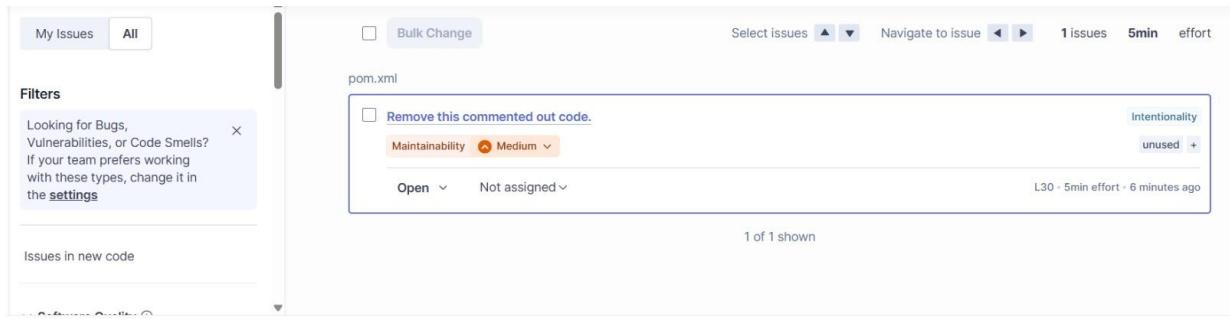


Figure 21: Analyse de maintenabilité

4.6. Correction de l'anomalie détectée

Suite à la détection de cette anomalie par SonarQube, l'ancienne API de génération aléatoire de NumPy a été remplacée par `numpy.random.Generator`, une solution plus moderne et robuste.

Cette amélioration contribue à renforcer la qualité du code, assurer la reproductibilité des traitements de données et limiter l'accumulation de dette technique.

A screenshot of the SonarQube code analysis interface for a file named 'etl.py'. The code editor shows a line of Python code: '# Génération des features data = { 'age': np.random.randint(18, 85, n_samples), ... }'. A red wavy underline highlights the 'np.random.randint' call. A tooltip above the code reads: 'Use a "numpy.random.Generator" here instead of this legacy function.' Below the code editor, there's a navigation bar with tabs for 'Where is the issue?', 'Why is this an issue?', 'How can I fix it?', 'Activity', and 'More info'. To the right, there's a sidebar with 'Software qualities impacted' (Maintainability, Medium), 'Code attribute' (Consistency | Not conventional), and links to 'Open in IDE' and 'See all issues in this file'. At the bottom, there's footer information including 'Community Build v25.11.0.114957' and various links.

Figure 22: Anomalie détectée

4.7. Résultat après correction

Après la correction et la relance de l'analyse, SonarQube n'a signalé **aucune anomalie restante**.

Le projet SmartPark présente désormais :

- **0 bug**
- **0 vulnérabilité**
- **0 code smell critique**
- **Quality Gate : Passed**

Ces résultats confirment que le projet respecte pleinement les critères de qualité logicielle définis.

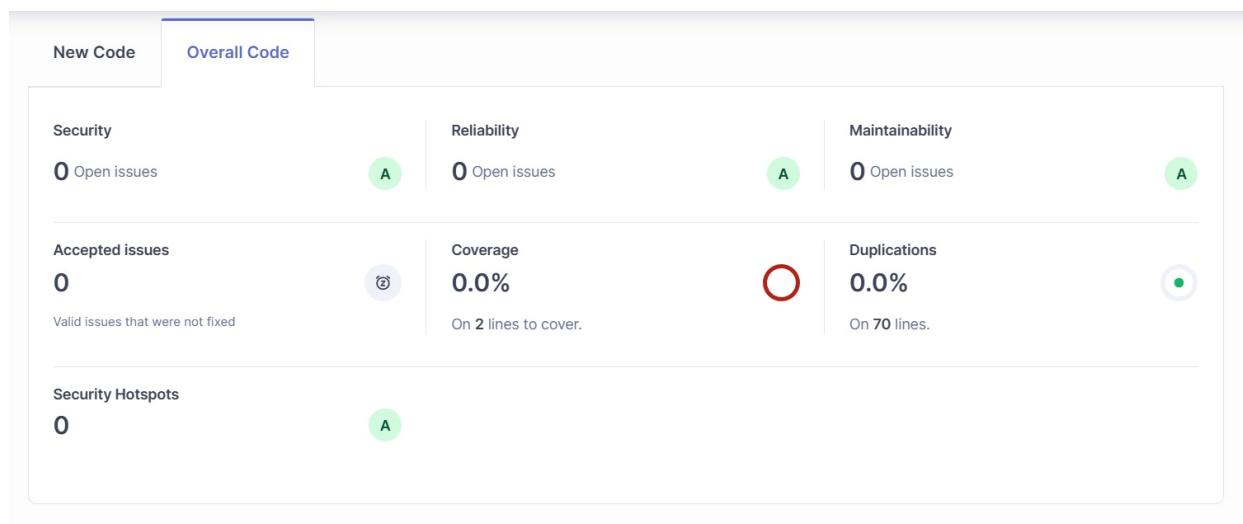


Figure 23: Résultat final

4.8. Bilan de la démarche qualité

L'intégration de SonarQube dans le projet Healthinsight a permis d'adopter une **démarche de qualité logicielle professionnelle**, similaire à celle utilisée dans le monde industriel. Grâce à cet outil, il a été possible d'identifier rapidement les anomalies, de les corriger efficacement et de valider la qualité globale du code avant la finalisation du projet.

Cette démarche contribue à améliorer la maintenabilité, la fiabilité et la pérennité du projet Healthinsight.

5. Conclusion

Ce projet de fin d'année avait pour objectif principal la conception et la réalisation de *HealthInsight*, une plateforme web intelligente dédiée à la prédition des risques de santé à partir de données statistiques issues de Statistique Canada. À travers ce travail, nous avons

Chapitre VI : Conclusion et Perspectives

cherché à répondre à une problématique actuelle liée à l'exploitation des données de santé et à leur valorisation par des techniques d'ingénierie des données et d'intelligence artificielle. Le projet a couvert l'ensemble du cycle de développement d'une application informatique, depuis l'analyse du besoin et l'étude de l'existant, en passant par la conception fonctionnelle et technique, jusqu'à l'implémentation et la mise en œuvre d'une solution opérationnelle. La mise en place d'un pipeline ETL, l'entraînement d'un modèle de machine learning, son intégration dans une API backend, ainsi que le développement d'une interface web ergonomique avec des visualisations interactives et 3D, ont permis d'aboutir à une solution cohérente, fonctionnelle et extensible. Les résultats obtenus démontrent la pertinence de l'approche adoptée et confirment que les outils d'analyse prédictive peuvent constituer une aide efficace à la compréhension et à l'anticipation des risques de santé. Au-delà de l'aspect technique, ce projet nous a permis de consolider nos compétences en génie logiciel, data science et développement web, tout en nous confrontant à des problématiques réelles liées à la qualité des données, à la sécurité et à l'expérience utilisateur. Des perspectives d'évolution peuvent être envisagées, notamment l'intégration de modèles prédictifs plus avancés, l'enrichissement des sources de données, ainsi que le déploiement de la plateforme dans un environnement de production, faisant de *HealthInsight* une base solide pour des travaux futurs à plus grande échelle.

Webographie

- [1] I. H. Witten, E. Frank, M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufmann, 2016.
- [2] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2017.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [4] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed., O'Reilly Media, 2019.
- [5] Statistics Canada, *Canadian Community Health Survey (CCHS)*, <https://www.statcan.gc.ca>, consulté en 2025.
- [6] J. Brownlee, *Machine Learning Mastery – Applied Machine Learning*, <https://machinelearningmastery.com>, consulté en 2025.
- [7] F. Chollet, *Deep Learning with Python*, Manning Publications, 2018.
- [8] S. Raschka, V. Mirjalili, *Python Machine Learning*, 3rd ed., Packt Publishing, 2019.
- [9] FastAPI Documentation, *FastAPI Framework*, <https://fastapi.tiangolo.com>, consulté en 2025.
- [10] React Documentation, *React – A JavaScript library for building user interfaces*, <https://react.dev>, consulté en 2025.
- [11] PostgreSQL Global Development Group, *PostgreSQL Documentation*, <https://www.postgresql.org/docs>, consulté en 2025.
- [12] World Health Organization (WHO), *Health statistics and information systems*, <https://www.who.int/data>, consulté en 2025.
- [13] OECD, *Health Data and Indicators*, <https://www.oecd.org/health>, consulté en 2025.
- [14] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.
- [15] M. Fowler, *UML Distilled: A Brief Guide to the Standard Object Modeling Language*, Addison-Wesley, 2004.