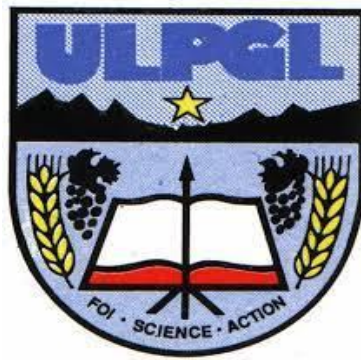


**UNIVERSITÉ LIBRE DES PAYS DES GRANDS LACS**

**Faculté des Sciences et Technologies Appliquées**

Département de Génie Electrique et Informatique



BP. 368 GOMA  
[www.ulpgl.net](http://www.ulpgl.net)

**Conception et réalisation d'une  
application web de diagnostic du  
diabète et de suivi du patient  
diabétique.**

Par : **AMANI HAMADA Bonheur.**

Travail présenté en vue de l'obtention du Diplôme de l'ingénieur  
Civil en Génie Électrique et Informatique.

Option : Génie Informatique

Directeur : Prof. Dr. **Olivier BARAKA MUSHAGE**

Encadreur : **Ir. Johnson KISAMBA**

**2022-2023**

# Epigraphe

---

*« La créativité est contagieuse, faites-la tourner »*

**Albert Einstein**

# Dédicace

---

A mes très chers parents, **HAMADA BALOLE** et **NSIMIRE ELISABETH**,

**AMANI HAMADA Bonheur**

## Remerciements

---

Nos remerciements les plus avoués s'adressent d'abord au Dieu d'ELIE, d'Isaac et de Branham pour son amour, sa grâce et sa protection qu'Il continue à nous offrir. Que toute gloire revienne à lui.

Nos remerciements s'adressent à l'ULPGL en générale et à nos autorités académiques en particulier pour l'éducation de qualité qu'ils nous ont offerte. Pour leurs efforts et leur tolérance nous disons merci du fond de nos cœurs.

Gratitude à notre cher directeur, Prof. Dr Olivier BARAKA MUSHAGE, et l'encadreur Ir. Johnson KISAMBA pour leur encouragement et les orientations qu'ils m'ont offertes.

Nous ne pouvons pas ne pas remercier le staff médical qui nous a aidés à mieux comprendre des éléments de leur domaine liés à notre travail. Nous pouvons citer Dr MUJARUGAMBA Bienvenu de l'hôpital New Hope Center, le Dr. MULIRO FARAJA et KALIONGO Nicolas qui est un étudiant en médecine.

A nos chers parents HAMADA Balole et NSIMIRE Elisabeth, nous disons merci pour le dévouement qu'ils ont manifesté à l'égard de notre scolarisation.

Nous ne manquerons pas de manifester notre gratitude envers l'Ir Siméon TABASIMWA pour ses orientations et envers nos camarades qui ont été pour nous des compagnons de lutte durant cette période de six ans d'université.

En fin, à tous ceux-là qui de près ou de loin ont contribué pour la réalisation de ce travail, que la profonde gratitude manifestée à l'égard de leurs efforts leur parvienne.

**AMANI HAMADA Bonheur**

## Résumé

---

L'apprentissage automatique a le potentiel d'améliorer le diagnostic du diabète en rendant le processus plus rapide et plus précis pour faire des traitements plus efficaces. Le diagnostic est une étape importante du traitement d'une maladie. Il permet de déterminer la cause de la maladie, la gravité de la maladie et le pronostic du patient. La posologie des médicaments est également influencée par le diagnostic. Le diabète étant souvent pris comme une maladie difficile à diagnostiquer vu sa complexité et difficile à guérir, les techniques des applications qui utilisent de l'intelligence artificielle seraient une très bonne aide au traitement de cette maladie et faire le suivi du patient minutieusement. C'est dans ce sens que ce travail se démarque et se montre d'une importance capitale pour tous les utilisateurs de l'application. Dans ce travail, nous avons employé l'apprentissage automatique par régression logistique afin d'exploiter les données médicales présentes dans un ensemble de données. Ce processus a permis de construire un modèle d'intelligence artificielle et d'en extraire une conclusion pour le diagnostic du diabète. Pour sa réalisation, nous avons d'abord fait une petite étude du diabète, comment faire son diagnostic, comment utiliser l'intelligence artificielle en médecin pour la cause qui nous été déjà fixée. Enfin, nous avons implémenté ce modèle et évalué le résultat trouvé. Comme résultat final de ce travail, nous avons un outil d'aide au diagnostic du diabète basé sur la régression logistique, qui peut aider à rendre plus rapide le diagnostic avec une précision de 97,6%. Cet outil permet aussi d'enregistrer des données liées au patient pour faciliter le suivi de ce dernier.

Mots clés : diabète, diagnostic, apprentissage automatique.

## Abstract

---

Machine learning through logistic regression method was used in this work to use medical data found in a dataset to build an artificial intelligence model, and output a result for the diagnosis of diabetes. Machine learning has the potential to improve the diagnosis of diabetes by making the process faster and more accurate, leading to more effective treatments. Diagnosis is an important step in the treatment of a disease. It allows for determining the cause of the disease, the severity of the disease, and the patient's prognosis. The dosage of medications is also influenced by the diagnosis. Diabetes is often seen as a difficult disease to diagnose due to its complexity and difficulty to cure. The techniques of applications that use artificial intelligence would be a great help in the treatment of this disease and in carefully monitoring the patient's data. This is how this work stands out and shows its importance to all users of the application. To achieve this, we first conducted a small study of diabetes, how to diagnose it, and how to use artificial intelligence in medicine for the cause that was already set for us. Then, we designed an intelligent diabetes diagnosis model and a patient monitoring system. Finally, we implemented and evaluated the results found. As a final result of this work, we have a diabetes diagnosis helping tool based on logistic regression, which can help to make the diagnosis faster with an accuracy of 97.6%. This tool also allows for recording patient-related data to facilitate their follow-up.

Keywords: diabetes, diagnosis, machine learning.

# Sommaire

---

<b>Epigraphe.....</b>	<b>i</b>
<b>Dédicace .....</b>	<b>ii</b>
<b>Remerciements .....</b>	<b>iii</b>
<b>Résumé.....</b>	<b>iv</b>
<b>Sommaire.....</b>	<b>vi</b>
<b>Sigles et abréviations.....</b>	<b>viii</b>
<b>Liste des figures.....</b>	<b>ix</b>
<b>Listes des tableaux .....</b>	<b>xi</b>
<b>Les Equations .....</b>	<b>xii</b>
<b>Introduction générale .....</b>	<b>1</b>
0.1. Contexte .....	1
0.2. Problématique et hypothèses .....	1
0.3. Objectifs .....	2
0.4. Méthodes et techniques de recherche utilisées.....	3
0.5. Délimitation du sujet .....	3
0.6. Subdivision du travail.....	4
<b>Chapitre 1 : Concepts théoriques .....</b>	<b>5</b>
1.1 Les applications mobiles .....	5
1.2 La télémédecine.....	9
1.3 Le diabète .....	14
1.4 Conclusion partielle.....	21
<b>Chapitre 2 : L'apprentissage automatique.....</b>	<b>22</b>
2.1. Introduction .....	22
2.2. Définitions de l'apprentissage automatique .....	22
2.3. Types d'apprentissage automatique .....	22
2.4. Quelques méthodes et techniques de l'apprentissage automatique.....	25
2.5. Application des méthodes d'Apprentissage automatique dans la Prédiction du Diabète. ....	37
2.6. Conclusion partielle.....	40
<b>Chapitre 3 : Conception et réalisation de l'application web.....</b>	<b>41</b>

3.1. Introduction .....	41
3.2. Présentation du système .....	41
3.3. Le modèle de l'intelligence artificiel .....	42
3.4. Conception fonctionnelle du système .....	44
3.5. Les interfaces du système.....	60
3.6. Conclusion partielle.....	67
<b>Conclusion générale.....</b>	<b>68</b>
<b>Références.....</b>	<b>A</b>
<b>Annexe.....</b>	<b>D</b>
Définition du modèle.....	D
Structure de dossiers.....	F

## Sigles et abréviations

---

<b>Admin</b>	: Administrateur
<b>API</b>	: Application Programming Interface
<b>Colab</b>	: Collaboratory DL
<b>FN</b>	: False Negative
<b>FP</b>	: False Positive
<b>TN</b>	: True Positive
<b>TP</b>	: True Negative
<b>IA ou AI</b>	: Intelligence Artificielle
<b>JS</b>	: JavaScript Material
<b>ML</b>	: Machine Learning
<b>OS</b>	: Operating System
<b>RDC</b>	: République Démocratique du Congo
<b>ULPGL</b>	: Université Libre des Pays des Grands Lacs
<b>IMC</b>	: Indice de masse corporelle
<b>RTTT</b>	: Réponse Thérapeutique au Traitement

## Liste des figures

Figure 1: illustration de la téléconsultation .....	10
Figure 2: illustration de la télé-expertise .....	11
Figure 3: croquis de la téléassistance médicale .....	11
Figure 4: Schéma de la régulation médicale .....	12
Figure 5: exemple de classification par la méthode des k-NN .....	26
Figure 6 Schéma d'un arbre de décision .....	28
Figure 7: Schéma d'un arbre de décision pour les attributs et les classes .....	29
Figure 8: Arbre de décision qui répond à la question si une personne est diabétique ou non.	31
Figure 9: Structure de l'algorithme Radom Forest .....	32
Figure 10: Courbe linéaire représentant la relation entre le risque de diabète et le taux de mortalité .....	34
Figure 11: Régression non-linéaire des données .....	35
Figure 12: Courbe polynomiale du second degré .....	36
Figure 13: courbe sigmoïde de régression linéaire .....	36
Figure 14: Architecture de l'outil SantePlus. ....	42
Figure 15: Diagramme d'exigences de l'application SantePlus. ....	45
Figure 16: Diagramme des cas d'utilisation de SantePlus. ....	46
Figure 17: Diagramme d'activité pour s'authentifier dans SantePlus .....	52
Figure 18: Diagramme d'activité pour créer un compte dans SantePlus .....	53
Figure 19 Diagramme d'activité pour rechercher un patient de SantePlus .....	53
Figure 20: Diagramme d'activité pour voir le profil .....	54
Figure 21: Diagramme d'activité pour s'envoyer des messages dans SantePlus .....	54
Figure 22: Diagramme d'activité pour se déconnecter de SantePlus .....	54
Figure 23: Modélisation de la base de données de l'application. ....	55
Figure 24: Visualisation de la base de données dans le compte superadmin. ....	58
Figure 25: visualisation des valeurs nulles du dataset .....	59
Figure 26: visualisation des valeurs aberrantes du dataset et leur élimination. ....	59
Figure 27: matrice de corrélation de différentes données du dataset .....	60
Figure 28: login page de l'application SantePlus .....	61
Figure 29: Page d'enregistrement des utilisateurs .....	62
Figure 30: admin dashboard ou tableau de bord de l'administrateur .....	63
Figure 31: Tableau de bord pour un patient .....	63

Figure 32:page de visualisation des médecins .....	64
Figure 33: Formulaire pour ajouter un médecin .....	64
Figure 34:page de diagnostic .....	65
Figure 35 : formulaire pour ajouter les données du jour.....	66
Figure 36:visualisation des données d'un patient.....	66
Figure 37: contacter un utilisateur .....	67

## Listes des tableaux

---

Tableau 1-1:Tableau comparatif des types d'applications en fonction des facteurs de choix ...	9
Tableau 2-1:Les résultats de la précision des algorithmes d'étude .....	38
Tableau 2-2:comparaison de la précision et de la justesse des différentes méthodes.....	39
Tableau 3-1 Affichage de cinq des données du dataset avant encodage .....	43
Tableau 3-2:Cas d'utilisation 001 - authentification .....	47
Tableau 3-3:Cas d'utilisation 002 – déconnexion .....	47
Tableau 3-4:Cas d'utilisation 003 – Créer un compte Médecin .....	48
Tableau 3-5:Cas d'utilisation 004 – Création compte patient .....	48
Tableau 3-6:Cas d'utilisation 005 – compléter et modifier son profil.....	49
Tableau 3-7:Cas d'utilisation 006 – Voir un patient.....	49
Tableau 3-8:Cas d'utilisation 007 – Voir la liste des patients .....	50
Tableau 3-9:Cas d'utilisation 008 – Faire un diagnostic rapide .....	50
Tableau 3-10:Cas d'utilisation 009 – Voir les résultats de diagnostic.....	51
Tableau 3-11: Cas d'utilisation 010 –Envoyer un Message. ....	51

# Les Equations

---

Équation 1: Valeur de l'indice de masse corporel .....	20
Équation 2: Fonction de mappage des entres et des sorties de l'apprentissage supervise .....	23
Équation 3: Théorème de Bayes .....	25
Équation 4: Théorème de Bayes pour le cas de la classification .....	26
Équation 5: La probabilité la plus élevée du théorème de Bayes .....	26
Équation 6: Recherche de la distance Euclidienne .....	27
Équation 7: Recherche de la distance de Minkowsky .....	27
Équation 8: Recherche de la distance de Manhattan .....	27
Équation 9: calcul de l'indice de Gini .....	29
Équation 10: calcul de l'entropie d'un attribut de l'ensemble .....	30
Équation 11: Calcul du gain d'information .....	30
Équation 12: fonction logistique .....	37
Équation 13: calcul de l'accuracy .....	37
Équation 14: Mesure de la précision du modèle .....	38
Équation 15: mesure du rappel ou la sensibilité .....	40
Équation 16 : calcul de F1-score .....	40

# Introduction générale

---

## 0.1. Contexte

Vu l'augmentation considérable du nombre de malades chroniques dans le monde et plus particulièrement les malades diabétiques, le manque d'infrastructures et moyens hospitaliers, la rareté des médecins spécialistes dans certains endroits, la télémédecine combinée à l'intelligence artificielle se montrent d'une importance capitale pour éradiquer ce genre de maladies.

Il n'y a pas très longtemps que l'intelligence artificielle (IA) a commencé à être mise en application dans le domaine des soins de santé moderne. De par ses techniques et ses algorithmes, l'IA intervient dans la conception des applications informatiques très performantes dans la prédiction et le diagnostic des maladies comme l'hypertension artérielle et le diabète.

En partant de l'hypothèse que les maladies sont plus facilement traitables lorsqu'elles sont détectées à un stade préliminaire, la prédiction des maladies joue un rôle crucial dans la mise en place de mesures de prévention, de diagnostic, et dans la prise de décisions relatives au traitement de maladies difficiles à soigner, telles que le diabète, qui est précisément le sujet de cette étude. [1]

## 0.2. Problématique et hypothèses

L'étape de diagnostic d'une maladie comme le diabète est très importante lors du dépistage. Pour assurer un traitement efficace et un suivi approprié du patient, il est essentiel de disposer d'une compréhension précise de la nature de la maladie. En d'autres termes, quelle est la maladie en question ? Et à quel stade est-elle ? Ces questions sont pertinentes et demandent une large précision dans un bref délai.

Un certain nombre des problèmes se posent à plusieurs niveaux : pour les personnels soignants et pour les patients.

De la part du médecin, la précision et le temps de diagnostic du diabète demeurent un problème vu la complexité et la dangerosité de cette maladie. De la part d'un patient, il doit se déplacer de son domicile vers l'hôpital fréquemment même lorsqu'il ne le fallait pas. Le suivi de cette maladie reste une tâche difficile surtout en utilisant des méthodes de traitement archaïques utilisées dans les hôpitaux.

Comment faire l'usage de la technologie pour subvenir aux besoins précédemment cités ?

Ainsi, nos questions de recherche peuvent être formulées de la manière suivante :

- Serait-ce possible de réduire le temps de diagnostic du diabète tout en ayant une haute précision ?
- Comment utiliser les nouvelles technologies comme un moyen pour faciliter le suivi du patient ?
- Un outil d'aide au diagnostic et de suivi du patient, serait-il le bienvenu dans les hôpitaux et agréé par les médecins surtout dans la zone Afrique et plus précisément en RDC ?

Les réponses anticipatives aux questions ci-dessus sont successivement les suivantes :

- La mise en œuvre d'une application web serait un moyen permettant, à partir d'un ensemble de données médicales, de détecter si la personne est atteinte du diabète et de prédire si elle est en phase de pré diabète. Cet outil aiderait à réduire le temps et avoir une haute précision.
- Une base de données permettrait de stocker les informations du malade, une fois intégrée dans notre système, celle-ci permettrait de voir l'évolution de la maladie. Il ne serait plus question de le faire comme à l'ancienne avec une paperasse qui causerait une perte d'informations et une mauvaise décision.
- Après une enquête faite auprès de deux médecins de l'hôpital New Hope center et un de La charité maternelle, cinq étudiants en médecine et dix patients diabétiques nous avons constaté que cette application web serait la bienvenue vu les avantages qu'elle apporte peu importe les contraintes. La précision de cette application lors de diagnostic et la facilité de faire le suivi du patient feraient que cette application soit un outil incontournable dans le traitement du diabète.

### **0.3.Objectifs**

#### 0.3.1. L'objectif général

Dans ce projet, nous envisageons de développer une application pratique et bénéfique qui, en utilisant des données médicales, sera capable de détecter la présence du diabète et de suivre l'évolution de cette condition chez une personne. Cet outil doit permettre l'automatisation des traitements, l'échange d'informations, la conservation des données ainsi que l'exécution rapide des tâches.

### 0.3.2. Les objectifs spécifiques

Pour arriver à l'objectif général, spécifiquement il nous faudra :

- Acquérir une connaissance suffisante sur les applications web.
- Comprendre d'une manière suffisante possible le diabète, les types de diabètes et leurs modes de diagnostic.
- Appliquer l'intelligence artificielle en télémédecine.
- Concevoir un outil intelligent de diagnostic.
- Concevoir un système de suivi du patient.
- Implémenter et évaluer le résultat trouvé.

## 0.4. Méthodes et techniques de recherche utilisées

### 0.4.1. Méthodes

- Méthode descriptive : elle consiste à appréhender notre sujet dans son ensemble et dans ses aspects particuliers.
- Méthode analytique : elle permet de faire une analyse systématique de données que nous possédons.

### 0.4.2. Techniques

- La technique documentaire : celle-ci nous a permis d'exploiter les ouvrages, les archives et les autres travaux de recherches scientifiques ayant trait à notre thème de recherche ;
- La technique expérimentale (simulation) : celle-ci nous a permis de faire des simulations pour plus de précision sur ce que nous présentons.

## 0.5. Délimitation du sujet

Ce travail va se limiter à l'implémentation d'un outil d'aide de diagnostic du diabète grâce aux méthodes et techniques de l'intelligence artificielle ainsi qu'à la conception d'un système de suivi du patient facile à utiliser pour tout le monde. Sur le plan temporel, cette application restera valide aussi longtemps que les techniques de diagnostic y attachées seront pertinentes pour la cause.

## 0.6.Subdivision du travail

Notre travail est divisé en deux parties : la **partie théorique** et la **partie développement de l'application**.

La partie théorique sera composée de deux chapitres :

- Le premier s'intitule « **Concepts théoriques de base** », il a pour but de présenter les concepts liés aux applications mobiles, d'aborder la maladie du diabète sur laquelle sera basée notre application ainsi que parler de la télémédecine.
- Le deuxième chapitre s'intitule « **L'apprentissage automatique** », il a pour but de connaître les différentes techniques et méthodes du Machine Learning pour bien faire le choix lors de la création du model de prédiction de la maladie.

La deuxième partie sera consacrée au cadre méthodologique de notre système. Cette partie sera composée du chapitre trois intitulé « **Conception et Réalisation de l'application** » par lequel nous saurons le fonctionnement de l'application et les différents liens qui sont entre les composants de notre application avant de passer à la réalisation de l'application en utilisant un nombre des technologies.

Pour clore, nous concluons ce travail en résumant les connaissances acquises durant la réalisation du projet, les problèmes rencontrés ainsi que quelques perspectives quant à l'avenir de notre application.

# Chapitre 1 : Concepts théoriques

---

## 1.1 Les applications mobiles

### 1.1.1. Introduction

De nos jours, les mobiles et les ordinateurs sont les appareils les plus utilisés pour communiquer, naviguer sur Internet et faire beaucoup des travaux dans les entreprises. Plus rien ne se fait comme à l'ancienne. L'utilisation intense de ces outils est à la base du développement des applications web et mobiles. Il est donc nécessaire de placer un mot sur ces notions avant d'entrer dans le vif du sujet.

### 1.1.2. Définitions

Une application mobile est un outil informatique conçu pour fonctionner sur un appareil mobile, tel qu'un Smartphone ou une tablette PC. Elle vise à fournir des services similaires à des applications de bureau accessibles sur ordinateurs. [2]

### 1.1.3. Types d'applications mobiles

Selon leurs modes de fonctionnement, les applications mobiles se répartissent en trois catégories : les applications natives (non connectée), les applications web (connectées) et les applications hybrides.

#### 1.1.3.1. Les applications natives [3]

Les applications natives sont des applications conçues pour s'exécuter sur une plate-forme ou un système d'exploitation particulier, tel que iOS, Android, Mac ou Windows. Elles sont téléchargées et installées directement sur l'appareil.

#### ➤ *Fonctionnement*

Le téléphone mobile ou la machine ordinateur lit tous les fichiers qu'il lui faut directement dans la mémoire de stockage, ou dans une base de données embarquée. Ce type d'application est destiné à remplir une fonction bien précise, sans pour autant avoir besoin d'évoluer dans le temps par leur contenu.

#### ➤ *Avantages*

- Contrôle accru sur l'expérience utilisateur globale
- Plus de fonctionnalités.
- Une plus grande sécurité

- Meilleure performance
- Optimisé pour chaque appareil spécifique
- L'application est toujours disponible, quel que soit l'état du réseau et où que l'on soit.

➤ ***Inconvénients***

- Le développement peut être plus coûteux.
- Nécessite plus d'entretien.
- Peut prendre plus de temps à développer.
- La multiplication des types d'appareils induit souvent des incompatibilités.

1.1.3.2. Les applications web (connectées)

Les applications web sont des logiciels applicatifs hébergés sur un serveur et accessibles via un navigateur web. Contrairement aux applications natives, l'utilisateur d'une application web peut utiliser n'importe quelle plateforme. Il lui suffit de se connecter à l'application à l'aide de son navigateur. [4]

➤ ***Avantages des applications web***

Les applications web sont devenues très populaires ces dernières années, car elles offrent de nombreux avantages :

- **Accessibilité** : Les applications web peuvent être utilisées sur n'importe quel appareil disposant d'un navigateur web, qu'il s'agisse d'un ordinateur, d'une tablette ou d'un Smartphone.
- **Fonctionnalités avancées** : Les applications web peuvent être très complexes et offrir des fonctionnalités avancées comparables à celles des logiciels traditionnels.
- **Mise à jour facile** : Les applications web sont mises à jour automatiquement par le serveur, ce qui signifie que les utilisateurs n'ont pas à se soucier de télécharger et d'installer des mises à jour.
- **Coût réduit** : Le développement et la maintenance d'une application web sont généralement moins coûteux que ceux d'un logiciel traditionnel.

Les applications web ont révolutionné la façon dont nous utilisons Internet. Elles nous permettent d'accéder à des services de haute qualité sans avoir à installer de logiciel sur notre ordinateur, tablette ou Smartphone.

Les applications web sont également utilisées pour le développement de nouvelles technologies, telles que l'intelligence artificielle et la réalité augmentée. Ces technologies sont de plus en plus utilisées dans les applications web pour offrir aux utilisateurs des expériences plus immersives et interactives.

➤ *Désavantages des applications Web [5]*

Les applications web présentent de nombreux avantages, mais elles ont également quelques inconvénients :

- **Performance** : Les applications web peuvent parfois être moins performantes que les applications natives, car elles doivent être chargées et exécutées depuis un serveur. Cela peut être un problème pour les applications qui nécessitent des performances élevées, telles que les jeux ou les applications de streaming vidéo.
- **Expérience utilisateur** : Les applications web peuvent parfois offrir une expérience utilisateur moins immersive que les applications natives, car elles sont conçues pour être utilisées dans un navigateur web. Cela peut être un problème pour les applications qui nécessitent une interaction utilisateur fluide et intuitive.
- **Compatibilité** : Les applications web sont généralement compatibles avec un large éventail d'appareils, mais elles peuvent ne pas être disponibles sur tous les appareils. Cela peut être un problème pour les applications qui doivent être accessibles à un large public.
- **Sécurité** : Les applications web peuvent être vulnérables aux attaques, car elles sont exécutées sur un serveur accessible au public. Cela peut être un problème pour les applications qui stockent des données sensibles, telles que des informations financières et/ou personnelles.
- Surtout au Congo, l'internet est d'une qualité déplorable et c'est une barrière pour l'utilisation des applications web.

#### 1.1.3.3. Les applications hybrides

Les applications hybrides sont des applications qui combinent les avantages des applications web et des applications natives. Elles sont généralement développées en utilisant un Framework hybride, comme React Native ou Ionic...

➤ *Avantages des applications hybrides*

Les applications hybrides présentent les avantages suivants :

- **Accessibilité** : Les applications hybrides peuvent être utilisées sur un large éventail d'appareils, y compris les ordinateurs, les tablettes et les smartphones.
- **Performance** : Les applications hybrides sont plus performantes que les applications web, car elles peuvent accéder aux fonctionnalités natives de l'appareil.
- **Expérience utilisateur** : Les applications hybrides peuvent offrir une expérience utilisateur comparable à celle des applications natives.

➤ *Inconvénients des applications hybrides*

Les applications hybrides présentent les inconvénients suivants :

- **Coût** : Le développement et la maintenance d'une application hybride peuvent être plus coûteux que ceux d'une application web, mais moins coûteux que ceux d'une application native.
- **Complexité** : Le développement d'une application hybride peut être plus complexe que le développement d'une application web ou native.
- **Flexibilité** : Les applications hybrides peuvent être moins flexibles que les applications natives, car elles sont limitées par le Framework hybride utilisé.

Les applications natives, web et hybrides sont trois types d'applications différents qui présentent chacun leurs propres avantages et inconvénients.

Les applications natives sont généralement le meilleur choix pour les applications qui doivent offrir une performance optimale et une expérience utilisateur immersive. Elles sont également le meilleur choix pour les applications qui nécessitent l'accès aux fonctionnalités natives de l'appareil, telles que la caméra, le microphone ou le GPS.

Les applications web sont généralement le meilleur choix pour les applications qui doivent être accessibles sur un large éventail d'appareils et qui n'ont pas besoin d'une expérience utilisateur immersive. Elles sont également le meilleur choix pour les applications qui doivent être mises à jour fréquemment.

Les applications hybrides sont un bon compromis entre les applications web et les applications natives. Elles offrent une certaine accessibilité et une certaine performance, mais elles ne sont pas aussi performantes ou immersives que les applications natives.

Le choix de la bonne solution dépend des besoins spécifiques de l'application. Il est important de prendre en compte les facteurs suivants lors du choix d'une solution :

**Performance, Expérience utilisateur, Accessibilité, Fonctionnalités et le Coût**

Le meilleur choix dépend des besoins spécifiques de l'application. Ici-bas un tableau comparatif des types d'applications en fonction des facteurs du choix d'un type d'applications.

**Tableau 1-1:Tableau comparatif des types d'applications en fonction des facteurs de choix**

<b>Caractéristiques</b>	<b>Application natives</b>	<b>Application Web</b>	<b>Application hybride</b>
Performance	Meilleure	Moins bonne	Entre les deux
Expérience utilisateur	Meilleure	Moins bonne	Entre les deux
Accessibilité	Moins bonne	Meilleure	Entre les deux
Fonctionnalités	Les plus complètes	Les moins complètes	Entre les deux
Coût	Le plus élevé	Le moins élevé	Entre les deux

## 1.2 La télémédecine

### 1.2.1 Introduction

La télémédecine est une pratique médicale à distance qui permet de fournir des soins médicaux à des patients situés à distance d'un professionnel de santé. Elle utilise des technologies de l'information et de la communication pour établir une connexion entre le patient et le professionnel de santé, qui peut être un médecin, un infirmier, un kinésithérapeute, etc. [2]

### 1.2.2 Définition

La télémédecine est définie comme "l'utilisation des technologies de l'information et de la communication pour fournir des soins médicaux à distance". Elle comprend cinq actes distincts : la téléconsultation, la télé expertise, la télésurveillance médicale, la téléassistance médicale, la régulation médicale. [6]

### 1.2.2.1 La téléconsultation

Pratique médicale à distance qui permet de fournir des soins médicaux à des patients situés à distance d'un professionnel de santé. [2] Elle permet aux patients de bénéficier de soins médicaux de qualité, même s'ils ne peuvent pas se rendre dans un cabinet ou un établissement de santé.

Voici une image illustrative d'un patient qui parle avec son médecin sans pouvoir se déplacer pour aller à l'hôpital.



**Figure 1: illustration de la téléconsultation [7]**

### 1.2.2.2 La télé expertise

La télé-expertise est une forme de télémédecine qui permet à un professionnel de santé (le demandeur) de solliciter à distance l'avis d'un ou de plusieurs professionnels de santé (les requis) en raison de leurs formations ou de leurs compétences particulières, sur la base des informations médicales liées à la prise en charge d'un patient. [7] C'est une pratique importante car elle permet la collaboration entre médecins.

La figure suivante montre bien une collaboration entre deux professionnels soignants qui sont dans des hôpitaux éloignés.

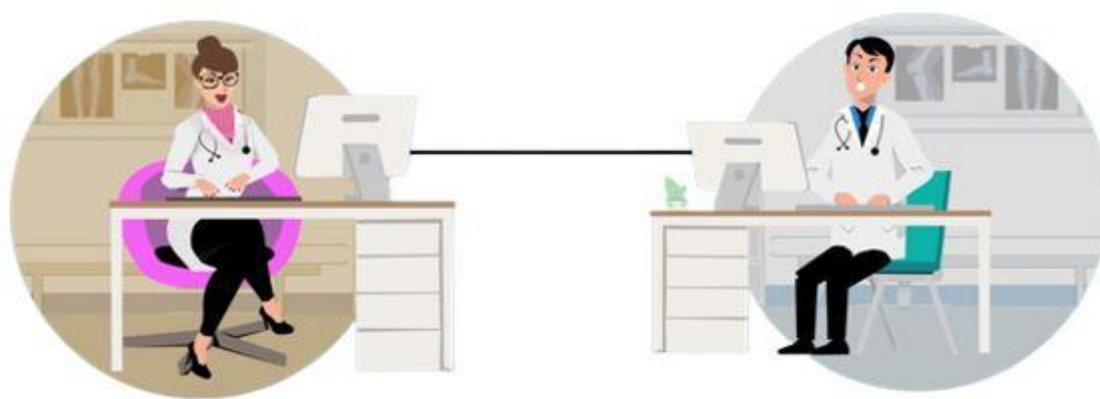


Figure 2: illustration de la télé-expertise [7]

### 1.2.2.3 La téléassistance médicale

La téléassistance médicale est un service qui permet aux personnes âgées, handicapées ou dépendantes, de bénéficier d'une assistance médicale à distance. Elle repose sur un dispositif d'alerte, généralement un médaillon ou une montre connectée, qui permet à la personne d'appeler une plateforme téléphonique en cas de problème [8].



Figure 3: croquis de la téléassistance médicale [6]

### 1.2.2.4 La régulation médicale

La régulation médicale est un acte médical pratiqué au téléphone par un médecin régulateur. Le médecin régulateur détermine et déclenche la réponse la mieux adaptée à l'état du patient, puis, si nécessaire, oriente le patient directement vers une unité d'hospitalisation appropriée. La régulation médicale a pour objectif de garantir l'accès aux soins urgents de qualité pour tous les patients, en optimisant l'utilisation des ressources hospitalières et en contribuant à une meilleure structuration de l'offre de soins sur un territoire de santé.

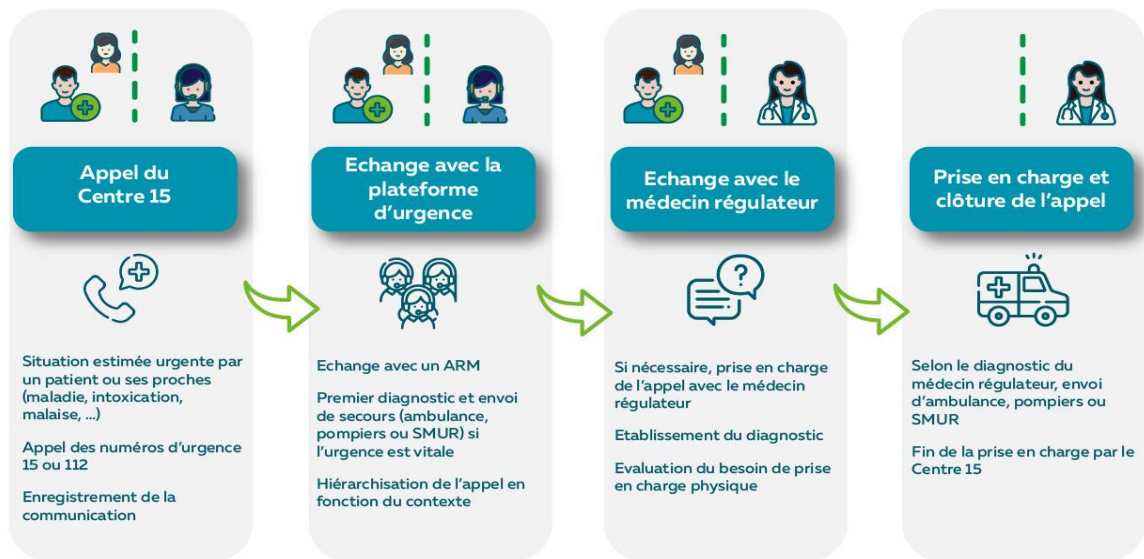


Figure 4:Schéma de la régulation médicale [7]

### 1.2.3 Avantages de la télémédecine. [6]

La télémédecine présente de nombreux avantages pour les patients, les professionnels de santé et le système de santé dans son ensemble.

➤ **Pour les patients :**

- Un meilleur accès aux soins, notamment dans les zones rurales ou reculées.
- Une réduction des délais d'attente pour obtenir un rendez-vous médical.
- Une meilleure prise en charge des patients chroniques.
- Une plus grande autonomie et une meilleure qualité de vie.

➤ **Pour les professionnels de santé :**

- Une meilleure coordination des soins entre différents professionnels de santé.
- Une meilleure utilisation des ressources.
- Une meilleure qualité des soins.

➤ **Pour le système de santé :**

- Une réduction des coûts de santé.
- Une amélioration de l'efficacité du système de santé.
- Une amélioration de la qualité des soins.

#### 1.2.4 Inconvénients de la télémédecine

La télémédecine présente de nombreux avantages ; Cependant, elle présente également certains inconvénients, notamment [9] :

- **La fracture numérique** : la télémédecine nécessite un accès à un ordinateur, à une tablette ou à un Smartphone, ainsi qu'à une connexion Internet. Cela peut être un obstacle pour les personnes âgées, les personnes à faible revenu ou celles qui vivent dans des zones rurales.
- **Les limites de l'examen physique** : la télémédecine ne permet pas au médecin d'effectuer un examen physique complet du patient. Cela peut être un problème pour certaines conditions médicales, telles que les maladies chroniques ou les blessures.
- **La confidentialité des données** : la télémédecine transmet des informations médicales sensibles par Internet. Il est important de s'assurer que les mesures de sécurité appropriées sont en place pour protéger ces données.

La liste n'est pas exhaustive, la nécessité d'une formation des professionnels de santé à la télémédecine et le risque de perte de confidentialité des données sont des éléments qui limitent l'usage de la télémédecine.

Voici quelques exemples concrets d'inconvénients de la télémédecine :

- Une personne âgée qui ne dispose pas d'un ordinateur ou d'une connexion Internet ne pourra pas accéder à la télémédecine.
- Un patient qui présente des signes d'une maladie grave, telle qu'une crise cardiaque ou un accident vasculaire cérébral, devra se rendre aux urgences, même s'il a consulté un médecin en ligne.
- Un patient qui partage ses informations médicales avec un médecin en ligne doit s'assurer que ses données sont sécurisées.

Quelques conseils pour limiter les inconvénients de la télémédecine :

- S'assurer de la disposition d'un outil informatique (un ordinateur, une tablette, un Smartphone) pouvant accéder à internet, et d'une connexion Internet fiable.
- Prudence lors du partage des informations médicales en ligne.

La télémédecine est une pratique médicale innovante qui a le potentiel de révolutionner l'accès aux soins. Elle offre de nombreux avantages pour les patients, les professionnels de santé et le système de santé dans son ensemble.

Il est important de prendre en compte ces inconvénients avant de choisir la télémédecine. Dans certains cas, la télémédecine peut être une solution efficace, mais elle ne convient pas à tous les patients ou à toutes les situations.

La télémédecine est en plein développement et son utilisation est amenée à se développer dans les années à venir. Elle est notamment encouragée par les pouvoirs publics, qui souhaitent améliorer l'accès aux soins dans les zones rurales ou reculées.

Les nouvelles technologies, telles que l'intelligence artificielle et la réalité virtuelle, ont le potentiel de faire évoluer la télémédecine et de la rendre encore plus accessible et efficace.

**Exemples d'utilisation de la télémédecine :** La télémédecine peut être utilisée dans de nombreux contextes différents. Tel est le cas de :

- La consultation de suivi pour un patient atteint d'une maladie chronique.
- Le suivi d'une grossesse.
- La prise en charge d'une urgence médicale.
- La consultation d'un patient handicapé.
- Etc.

Donnons quelques exemples d'utilisation de la télémédecine :

- Un patient atteint de diabète peut utiliser une application de télésurveillance pour suivre son taux de glycémie.
- Un patient atteint d'une maladie chronique peut consulter un médecin à distance pour une consultation de suivi.
- Une femme enceinte peut suivre sa grossesse à distance avec l'aide d'une sage-femme.
- Un patient victime d'un accident peut être pris en charge par un médecin régulateur à distance.
- Etc.

## 1.3 Le diabète

### 1.3.1 Définition :

Le diabète est une maladie chronique qui se caractérise par une hyperglycémie, c'est-à-dire une augmentation du taux de sucre dans le sang. [10]

Le sucre est connu comme la principale source d'énergie de l'organisme, et il est apporté par l'alimentation. L'insuline est une hormone produite par le pancréas qui permet au

sucres de pénétrer dans les cellules pour être utilisés. Lorsque le pancréas ne produit pas suffisamment d'insuline ou que l'organisme ne l'utilise pas correctement, le sucre s'accumule dans le sang. Cela peut entraîner des complications graves, notamment des maladies cardiaques, des accidents vasculaires cérébraux, des lésions rénales et des problèmes de vision.

### 1.3.2 Les Types de Diabète [10]

Il existe deux principaux types de diabète : le type 1 et le type 2.

#### 1.3.2.1 Diabète de type 1

Le diabète de type 1 est un diabète touchant les personnes jeunes. La cause exacte reste inconnue mais elle est connue comme une pathologie auto-immune qui cause la destruction des cellules bêta du pancréas qui sécrètent normalement l'insuline. Certains pensent que c'est une maladie génétique qui attaque certaines cellules du pancréas. Le risque grimpe à 70 %. Dans le diabète de type 1, le risque se situe entre 4 et 8%, plus précisément 8 % si le père est diabétique, 4 % si c'est la mère (mais 30 % si les deux parents le sont). L'unique traitement actuellement est l'apport d'insuline : soit sous forme d'injections, soit avec une pompe à insuline. Les signes et symptômes du diabète de type 1 apparaissent généralement, en quelques semaines ou quelques mois. Ils peuvent inclure :

- Une soif intense
- Une miction fréquente
- Une perte de poids inexpliquée
- Une fatigue excessive
- Une vision troublée
- Des infections fréquentes
- Etc.

#### 1.3.2.2 Diabète de type 2

Le diabète de type 2 apparaît surtout chez des personnes en surpoids après 40 ans. Il est dû à une résistance du corps à l'insuline, cette hormone ne parvenant plus à faire rentrer le sucre dans les cellules. D'autres sources sont la génétique, l'obésité et le manque d'activité. Lorsque l'un des deux parents est diabétique de type 2, le risque de transmission à la descendance est de l'ordre de 40 % et 70% si les deux parents sont atteints.

Le diabète de type 2 étant une maladie évolutive, après l'augmentation progressive des antidiabétiques, des injections d'insuline seront proposées en complément au patient lorsque la carence en insuline sera trop importante. La résistance à l'insuline peut être due à un certain nombre de facteurs, notamment :

- L'obésité
- Le manque d'exercice
- Un régime alimentaire malsain
- Un âge avancé
- Un antécédent familial de diabète

Les signes et symptômes du diabète de type 2 peuvent être subtils et ne se manifestent qu'après de nombreuses années. Ils peuvent inclure :

- Une soif intense
- Une miction fréquente
- Une fatigue excessive
- Une vision trouble
- Des infections fréquentes

Le diabète de type 2 peut être contrôlé avec un traitement adapté. Le traitement vise à maintenir la glycémie à un niveau normal. Il peut inclure des changements alimentaires, de l'exercice physique, des médicaments oraux ou des injections d'insuline.

### 1.3.2.3 Autres types de diabète

En plus du diabète de type 1 et de type 2, il existe d'autres types de diabète, notamment :

- **Diabète gestationnel** : Il survient pendant la grossesse et disparaît généralement après l'accouchement. Cependant, les femmes qui ont eu un diabète gestationnel ont un risque accru de développer un diabète de type 2 plus tard dans la vie.
- **Diabète secondaire** : Il est causé par certaines maladies ou certains médicaments. Les causes les plus courantes de diabète secondaire sont :
  - Les maladies du pancréas, telles que la pancréatite ou la fibrose kystique
  - Les maladies des glandes surrénales
  - Les maladies de la thyroïde

- Certains médicaments, tels que les stéroïdes ou les médicaments anticancéreux

### 1.3.3 Les complications du diabète

Les complications du diabète sont des problèmes de santé qui peuvent survenir lorsque le diabète n'est pas contrôlé. Ces complications peuvent être graves et même mortelles. Les complications du diabète sont divisées en deux catégories [11]:

- Les complications aiguës surviennent rapidement et peuvent être mortelles si elles ne sont pas traitées. Les complications aiguës les plus courantes sont l'hypoglycémie et l'acidocétose diabétique.
- Les complications chroniques surviennent sur une longue période et peuvent être irréversibles. Les complications chroniques les plus courantes sont les maladies cardiaques, les accidents vasculaires cérébraux, les lésions rénales, les problèmes de vision et les problèmes de pied.

#### **Complications aiguës du diabète :**

- Hypoglycémie : C'est une condition dans laquelle le taux de sucre dans le sang est trop bas. Les symptômes de l'hypoglycémie peuvent inclure des sueurs, des tremblements, des maux de tête, des vertiges et des difficultés à se concentrer. Dans les cas graves, l'hypoglycémie peut entraîner une perte de conscience ou un coma.
- Acidocétose diabétique : C'est une condition dans laquelle le corps ne peut pas produire suffisamment d'insuline pour utiliser le glucose comme source d'énergie. L'acide cétonique s'accumule alors dans le sang, ce qui peut entraîner une acidose. Les symptômes de l'acidocétose diabétique peuvent inclure des nausées, des vomissements, une douleur abdominale, une respiration rapide et une fatigue. Dans les cas graves, l'acidocétose diabétique peut entraîner un coma ou la mort.

#### **Complications chroniques du diabète :**

- Maladies cardiaques : Les personnes diabétiques ont un risque plus élevé de développer des maladies cardiaques, notamment une crise cardiaque et un accident vasculaire cérébral.
- Accidents vasculaires cérébraux : Les personnes diabétiques ont un risque plus élevé de subir un accident vasculaire cérébral.
- Lésions rénales : Les personnes diabétiques ont un risque plus élevé de développer une insuffisance rénale.

- Problèmes de vision : Les personnes diabétiques ont un risque plus élevé de développer des problèmes de vision, notamment une rétinopathie diabétique, qui peut entraîner une cécité.
- Problèmes de pied : Les personnes diabétiques ont un risque plus élevé de développer des infections, des plaies et des amputations des pieds.

#### 1.3.4 Relation entre le Diabète et la Télémédecine

La télémédecine peut être un outil précieux pour les personnes atteintes de diabète. Elle peut leur permettre d'accéder aux soins dont elles ont besoin, même si elles vivent dans des zones rurales ou isolées. Elle peut également leur permettre de recevoir des soins de manière plus pratique et flexible, ce qui peut être particulièrement utile pour les personnes qui travaillent ou qui ont des responsabilités familiales.

##### 1.3.4.1 Avantages de la télémédecine pour les diabétiques [10]

La télémédecine offre un certain nombre d'avantages pour les personnes atteintes de diabète, notamment :

- **Un meilleur accès aux soins :** La télémédecine peut permettre aux personnes atteintes de diabète d'accéder aux soins dont elles ont besoin, même si elles vivent dans des zones rurales ou isolées.
- **Une plus grande commodité :** La télémédecine peut permettre aux personnes atteintes de diabète de recevoir des soins de manière plus pratique et flexible.
- **Une meilleure coordination des soins :** La télémédecine peut faciliter la coordination des soins entre les professionnels de la santé, ce qui peut améliorer la qualité des soins

##### 1.3.4.2 Les applications de la télémédecine pour le diabète

La télémédecine peut être utilisée pour une variété d'applications dans le cadre du diabète, notamment :

- **Consultations :** La télémédecine peut être utilisée pour fournir des consultations à des patients atteints de diabète. Cela peut être utile pour les patients qui ont besoin de conseils ou d'ajustements de leur traitement.
- **Diagnostics :** La télémédecine peut être utilisée pour effectuer des diagnostics à distance, tels que la mesure de la glycémie ou l'analyse d'un pied à risque.

- **Traitements** : La télémédecine peut être utilisée pour fournir des traitements à distance, tels que la formation à l'auto surveillance de la glycémie ou la télésurveillance.
- **Suivi** : La télémédecine peut être utilisée pour suivre les patients atteints de diabète à distance. Cela peut aider à identifier les problèmes potentiels plus tôt et à améliorer l'observance du traitement.

#### 1.3.4.3 Les limites de la télémédecine pour le diabète

La télémédecine n'est pas sans limites. Il est important de noter que la télémédecine ne peut pas remplacer une consultation en personne avec un professionnel de la santé. Dans certains cas, une consultation en personne peut être nécessaire pour effectuer des tests ou des traitements qui ne peuvent pas être réalisés à distance.

#### 1.3.5 Diagnostic de Diabétique.

Le diagnostic du diabète est basé sur une combinaison de facteurs, notamment les symptômes, les antécédents familiaux, les facteurs de risque et les tests de diagnostic.

##### 1.3.5.1 Les symptômes.

Le diabète peut provoquer une variété de symptômes, notamment :

- Une soif intense
- Une miction fréquente
- Une perte de poids inexpliquée
- Une fatigue excessive
- Une vision trouble
- Des infections fréquentes

##### 1.3.5.2 Les antécédents familiaux

Les antécédents familiaux sont un facteur de risque important pour le diabète. Si quelqu'un a un parent ou un frère ou une sœur atteinte de diabète, la personne a un risque plus élevé de développer la maladie. Le risque de diabète est encore plus élevé si cette personne a des antécédents familiaux de diabète de type 1 ou de diabète de type 2.

### 1.3.5.3 Les facteurs de risque

Les facteurs de risque du diabète comprennent :

- **Le surpoids et l'obésité :** L'indice de masse corporelle (IMC) est un indicateur du surpoids et de l'obésité. Il est calculé en divisant le poids par la taille au carré. Les personnes obèses ou en surpoids ont un risque plus élevé de développer un diabète de type 2.

$$IMC = \text{poids(en kilogrammes)} \div \text{taille(en mètre carrée)}$$

#### Équation 1: Valeur de l'indice de masse corporel

L'IMC est interprété comme suit :

- IMC inférieur à 18,5 : poids insuffisant
  - IMC compris entre 18,5 et 24,9 : poids normal
  - IMC compris entre 25 et 29,9 : surpoids
  - IMC supérieur ou égal à 30 : obésité
- **La sédentarité :** La sédentarité entraîne une accumulation de graisse abdominale, qui est un facteur de risque important pour le diabète de type 2. La graisse abdominale sécrète des hormones qui peuvent perturber la régulation de la glycémie.
  - **Un âge avancé et le genre** sont des facteurs importants à prendre en compte lors du diagnostic du diabète. Le diabète de type 1 est plus fréquent chez les garçons que chez les filles. Le diabète de type 2 est plus fréquent chez les femmes que chez les hommes, en particulier après la ménopause. Le diabète de type 1 est généralement diagnostiqué chez les enfants et les adolescents, mais il peut se manifester à tout âge. Le diabète de type 2 est généralement diagnostiqué chez les adultes d'âge moyen ou plus âgés, mais il peut également se manifester chez les enfants et les adolescents.

### 1.3.5.4 Les tests de diagnostic. [10]

Les tests de diagnostic du diabète sont des tests qui mesurent la glycémie, qui est le taux de sucre dans le sang. Les tests de diagnostic du diabète comprennent :

- **Glycémie à jeun :** La glycémie à jeun est le test de diagnostic du diabète le plus courant. Ce test mesure le taux de sucre dans le sang après avoir jeûné pendant au moins 8 heures. Une glycémie à jeun supérieure ou égale à 1,26 g/L (7 mmol/L) indique un diabète.

- **Glycémie post-prandiale** : La glycémie post-prandiale est un test qui mesure le taux de sucre dans le sang 2 heures après avoir mangé. Une glycémie post-prandiale supérieure ou égale à 2 g/L (11,1 mmol/L) indique un diabète.
- **Test d'hémoglobine glyquée (HbA1c)** : L'HbA1c est une mesure de la glycémie moyenne sur une période de 2 à 3 mois. Une HbA1c supérieure ou égale à 6,5 % indique un diabète.
- **Test oral de tolérance au glucose (OGTT)** : L'OGTT est un test qui mesure la façon dont le corps réagit à une dose de glucose. Ce test est généralement utilisé pour diagnostiquer le diabète gestationnel, qui est un type de diabète qui survient pendant la grossesse.
- **Test de la glycémie capillaire** : Le test de la glycémie capillaire est un test rapide qui mesure le taux de sucre dans le sang à l'aide d'une petite piqûre au doigt. Ce test peut être utilisé pour diagnostiquer le diabète ou pour surveiller la glycémie chez les personnes atteintes de diabète.
- **Test de la glycémie interstitielle** : Le test de la glycémie interstitielle mesure le taux de sucre dans le liquide qui entoure les cellules. Ce test peut être utilisé pour surveiller la glycémie chez les personnes atteintes de diabète de type 1.

#### 1.4 Conclusion partielle

Dans ce chapitre, nous avons abordé les concepts fondamentaux liés aux applications mobiles et à la télémédecine. Ces éléments seront essentiels pour diagnostiquer le diabète et assurer le suivi d'un patient ayant été testé positif. Une introduction au diabète a été fournie afin de permettre la compréhension de cette maladie, étant la pierre angulaire de notre étude. La connaissance de cette pathologie est cruciale avant d'aborder le sujet de l'apprentissage automatique. Ce dernier constituera le deuxième chapitre de notre travail, visant à explorer comment l'intelligence artificielle peut être exploitée dans le traitement de cette maladie.

## Chapitre 2 : L'apprentissage automatique

---

### 2.1. Introduction

Ces dernières années, des améliorations considérables ont été observées dans les domaines d'applications tels que la vision par ordinateur, la reconnaissance de la parole et d'image, la traduction automatique et la télémédecine pour le diagnostic et le suivi de certaines maladies. Ces progrès ont été essentiellement stimulés par les récentes avancées en apprentissage automatique.

Dans ce chapitre, nous présentons un ensemble de généralités et de concepts relatifs à l'apprentissage automatique. En premier lieu nous donnons quelques définitions de l'apprentissage automatique et nous présentons leurs différents types, méthodes et Techniques.

Après cela, nous présentons l'application des méthodes et techniques de l'apprentissage automatique dans la prédiction du diabète.

### 2.2. Définitions de l'apprentissage automatique

L'apprentissage automatique regroupe tellement de concepts différents et variés qu'il est difficile d'en donner une définition unique. Pour cela, nous donnons deux définitions qui nous ont intéressés.

**Définition 1** : L'apprentissage automatique est un domaine de l'intelligence artificielle qui fait référence au développement, à l'analyse et à l'implémentation de méthodes automatisables permettant à une machine d'évoluer grâce à un processus d'apprentissage, et ainsi, remplir des tâches qu'il est difficile ou impossible de satisfaire par des moyens algorithmiques plus classiques [12].

**Définition 2** : L'apprentissage automatique est une discipline scientifique utilisée en intelligence artificielle. Il s'agit d'algorithmes (procédures traduites en langages informatiques) qui analysent un ensemble de données afin de déduire des règles constituant des nouvelles connaissances et permettant d'analyser des nouvelles situations [13].

### 2.3. Types d'apprentissage automatique

Il existe plusieurs types d'apprentissage automatique, qui se distinguent essentiellement par leur objectif. Parmi les types d'apprentissage automatique, les plus largement adoptés sont :

l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage par renforcement ainsi que l'apprentissage semi-supervisé.

### 2.3.1. L'apprentissage supervisé

Ce type d'apprentissage est principalement utilisé dans les applications en temps réel et l'approche pratique. Dans cet apprentissage, le modèle essaie d'apprendre des informations de l'expérience précédente qui lui sont données ; ces informations sont toutes étiquetées [14].

Mathématiquement, l'entrée ( $x$ ) et la sortie ( $y$ ) sont déterminées par un algorithme qui donne la fonction de mappage ( $f$ ) de l'entrée à la sortie comme indiqué ci-dessous :

$$y = f(x)$$

#### Équation 2: Fonction de mappage des entres et des sorties de l'apprentissage supervise

Les problèmes de classification et de régression sont les deux principaux domaines où l'apprentissage supervisé est utile :

- **La classification** : elle consiste à identifier les classes d'appartenance de nouveaux objets à partir d'exemples antérieurs connus, la variable à prédire peut donc prendre des valeurs discrètes appelées classes [15] (exemple : bon, moyen, mauvais).
- **La régression** : utilisée lorsqu'il s'agit de prédire une variable continue, qui peut donc prendre un nombre infini de valeurs (exemple : température, gains monétaires, etc.). Par exemple, pour une expérience de biochimie, on pourrait vouloir prédire le taux de résistance d'un organisme en fonction des taux de différentes substances qui lui sont administrées [15].

Il existe des dizaines d'algorithmes développés pour l'apprentissage supervisé, et chacun d'eux utilise diverses méthodologies pour prédire la valeur de la sortie.

La construction et la génération d'un modèle en apprentissage supervisé passe par deux phases : la phase d'apprentissage (d'entraînement) et la phase de test.

#### Phase d'apprentissage (d'entraînement)

Dans cette phase, l'algorithme d'apprentissage reçoit en entrée des exemples d'apprentissage (les documents d'entraînement) étiquetés et produit un modèle de prédiction le plus performant possible, c'est-à-dire le modèle qui produit le moins d'erreurs.

## Phase de test

Dans cette phase, le modèle obtenu lors de la phase d'apprentissage doit être capable de prédire l'étiquette d'un nouvel exemple en fonction des valeurs d'entrées.

### 2.3.2. L'apprentissage non-supervisé

Dans l'apprentissage non supervisé, les algorithmes essaient de découvrir un modèle ou une caractéristique unique eux-mêmes à l'insu de l'expérience antérieure. Mathématiquement, le modèle a une entrée ( $x$ ) mais n'a pas de correspondance de sortie. Ce type d'apprentissage est dit non supervisé car elle ne nécessite pas l'étiquetage des données de la base d'apprentissage et la machine ou le système trouve lui-même la réponse de la sortie correcte. Les algorithmes basés sur l'apprentissage non supervisé sont utilisés dans les problèmes liés à l'association et regroupement [14].

Aucun expert n'est disponible. Il vise à concevoir un modèle structurant l'information. La différence ici est que les comportements (ou catégories ou encore les classes) des données d'apprentissage ne sont pas connus, c'est ce que l'on cherche à trouver.

### 2.3.3. L'apprentissage par renforcement

L'apprentissage par renforcement fait référence à une classe de problèmes d'apprentissage automatique, dont le but est d'apprendre, à partir d'expériences successives, ce qu'il convient de faire de façon à trouver la meilleure solution [14].

Dans un tel problème, on dit qu'un « agent » (l'algorithme, au sens du code et des variables qu'il utilise) interagit avec « l'environnement » pour trouver la solution optimale. L'apprentissage par renforcement diffère fondamentalement des problèmes supervisés et non supervisés par ce côté interactif et itératif : l'agent essaie plusieurs solutions (on parle « d'exploration »), observe la réaction de l'environnement et adapte son comportement (les variables) pour trouver la meilleure stratégie (il « exploite » le résultat de ses explorations). Un des concepts clés de ce type de problèmes est l'équilibre entre ces phases d'exploration et d'exploitation.

### 2.3.4. Apprentissage semi supervisé

L'apprentissage semi-supervisé est une classe de techniques d'apprentissage automatique qui utilise un ensemble de données étiquetées et non étiquetées. Il se situe ainsi entre l'apprentissage supervisé qui n'utilise que des données étiquetées et l'apprentissage non supervisé qui n'utilise que des données non étiquetées. Il a été démontré que l'utilisation de

données non étiquetées, en combinaison avec des données étiquetées, permet d'améliorer significativement la qualité de l'apprentissage [14]. Un autre intérêt provient du fait que l'étiquetage de données nécessite souvent l'intervention d'un utilisateur humain. Lorsque les jeux de données deviennent très grands, cette opération peut s'avérer fastidieuse.

## 2.4. Quelques méthodes et techniques de l'apprentissage automatique.

Dans cette partie du travail, nous parlons des méthodes et techniques qui sont utilisées par les différents types d'apprentissage qui ont été le sujet la partie précédente. Avant de continuer, il est important de comprendre c'est quoi une méthode et c'est quoi une technique.

Une méthode est une catégorie générale de tâches que les machines peuvent apprendre à effectuer. C'est le cas de la **classification**, la **régression** et le **clustering**. Mais, une technique c'est un modèle ou un algorithme spécifique qui peut être utilisé pour effectuer une tâche.

### 2.4.1. Les méthodes de classification

Comme dit précédemment, les méthodes de classification s'appliquent lorsque l'ensemble des valeurs résultats est discret. Ceci revient à attribuer une classe (aussi appelée étiquette ou label) pour chaque valeur d'entrée. Les méthodes de classification peuvent être basées sur des hypothèses probabilistes (exemple, naïf bayésien), des notions de proximité (exemple, k plus proches voisins) ou des recherches dans des espaces d'hypothèses (exemple, arbres de décisions) [16]. Cherchons à avoir une idée sur quelques-unes de ces méthodes.

#### a) Naïf Bayésien

La classification naïve bayésienne repose sur l'hypothèse que toutes les caractéristiques sont conditionnellement indépendantes les unes des autres. Cette méthode est basée sur le théorème de Bayes qui calcule la probabilité d'un événement à l'aide de la connaissance au préalable des conditions connexes. Ce théorème a été découvert par un statisticien anglais, Thomas Bayes, au 18ème siècle mais il n'a jamais publié son travail. Après son décès, ses notes ont été éditées et publiées par le mathématicien Richard Price [16]. Le théorème est donné par la formule suivante :

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

#### Équation 3:Théorème de Bayes

- $A$  et  $B$  sont des événements.
- $P(A)$  est la probabilité d'observer l'événement  $A$

- $P(B)$  est la probabilité d'observer l'événement  $B$ .
- $P(A|B)$  est la probabilité conditionnelle d'observer  $A$ , sachant qu'un autre événement  $B$  de probabilité non nulle s'est réalisé.

Dans un problème de classification, notre tâche est de trouver l'étiquette la plus probable  $A$ , étant donné les caractéristiques  $B$ , le théorème de Bayes devient :

$$P(y|X_1 \dots X_n) = \frac{P(x_1, \dots, x_n|y)P(y)}{P(X_1 \dots X_n)}$$

#### Équation 4: Théorème de Bayes pour le cas de la classification

Où  $n$  représente le nombre de caractéristiques,  $y$  est l'événement qu'on cherche à classer. Par conséquent, en tenant compte de l'hypothèse d'indépendance, Bayes prédit la classe qui constitue la probabilité la plus élevée par l'équation 5.

$$\hat{y} = \arg \max_{y \in Y} P(y|x) = \operatorname{argmax} \frac{P(x|y)P(y)}{P(x)}$$

#### Équation 5: La probabilité la plus élevée du théorème de Bayes

b) K-Plus proches voisins

Parmi les algorithmes d'apprentissage automatique les plus basiques, le k-plus proche voisin, souvent abrégé en k-NN où  $k$  est un entier positif. En Data Science, cet algorithme est largement utilisé pour les problèmes de classification des données. Mais avant de se lancer dans cette méthode, il faut savoir que les calculs peuvent s'avérer très coûteuses en temps de calcul, ainsi, les données doivent être prétraitées. Cette méthode peut être également utilisée dans les problèmes de régression. Prenons par exemple le problème de classification suivant : Dans le diagramme ci-dessous, il y a des objets ronds verts et des objets carrés bleus. Ceux-ci appartiennent à deux classes différentes : la classe des ronds et la classe des carrés. Lorsqu'un nouvel objet est inséré dans l'espace - dans ce cas, un cercle rouge - nous voulons que l'algorithme d'apprentissage automatique classe le cercle dans une certaine classe.

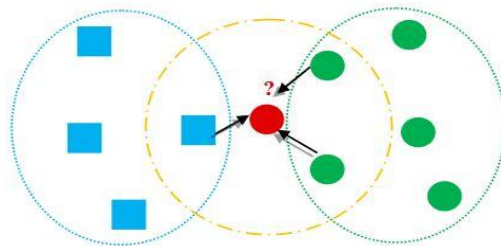


Figure 5: exemple de classification par la méthode des k-NN

Si on choisit  $k = 3$ , l'algorithme cherche les trois plus proches voisins du cercle rouge pour pouvoir le classer soit dans la classe des cercles, soit dans la classe des carrés. Dans ce cas, les trois plus proches voisins du cercle rouge sont un carré et deux cercles. Par conséquent, l'algorithme classera la sphère dans la classe des cercles.

Dans la méthode de k-NN, le résultat est l'appartenance à une classe. L'algorithme stocke tous les cas disponibles et classe tout nouvel objet en vérifiant ses k-plus proches voisins. Ensuite, l'objet est affecté à la classe qui contient le plus d'individus parmi les candidats trouvés. [17]

Cette méthode utilise principalement deux paramètres : une fonction de similarité pour comparer les individus dans l'espace de caractéristiques et le nombre k qui décide combien de voisins influencent la classification. Pour tester la similarité entre deux vecteurs, la distance est utilisée. Elle permet de mesurer le degré de différence entre deux vecteurs. Il existe plusieurs types de distance parmi lesquels on trouve : [17]

- La distance Euclidienne :  $\mathbf{d}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}_i)^2}$  x et y sont des vecteurs.

#### Équation 6:Recherche de la distance Euclidienne

- La distance de Minkowsky:  $\mathbf{d}(\mathbf{x}, \mathbf{y}) = (\sum_{i=1}^n |\mathbf{x}_i - \mathbf{y}_i|^p)^{1/p}$ . x et y sont des vecteurs.

#### Équation 7:Recherche de la distance de Minkowsky

- La distance de Manhattan :  $\mathbf{d}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |\mathbf{x}_i - \mathbf{y}_i|$  avec x et y des vecteurs

#### Équation 8:Recherche de la distance de Manhattan

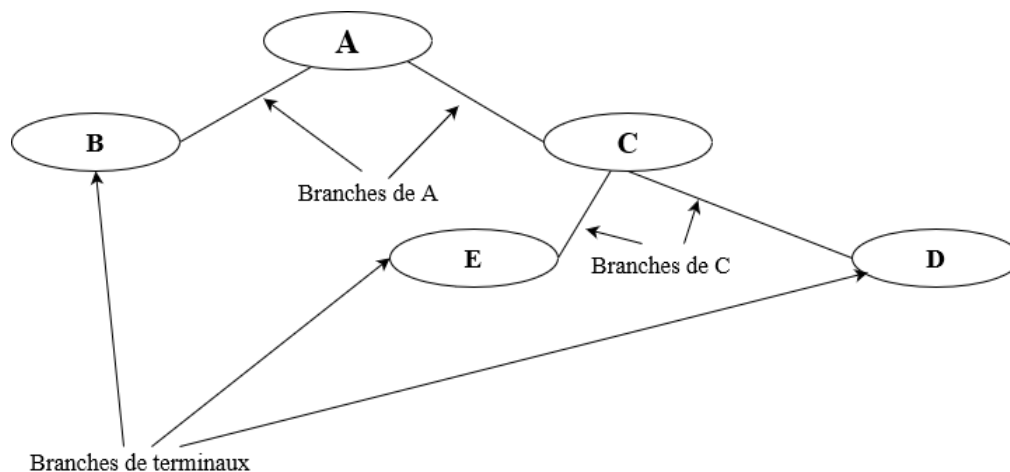
- **Avantages de la méthode des k plus proches voisins :**
  - L'algorithme K-NN est robuste envers des données bruitées.
  - La méthode des k plus proches voisins est efficace si les données sont larges et incomplètes.
  - Cette méthode est l'une des plus simples de tous les algorithmes d'apprentissage automatique.
- **Inconvénients de la méthode des k plus proches voisins :**
  - Le besoin de déterminer la valeur du nombre des plus proches voisins (le paramètre k).
  - Le temps de prédiction est très long puisqu'on doit calculer la distance de tous les exemples.

- Cette méthode est gourmande en espace mémoire car elle utilise une grande capacité de stockage pour le traitement des corpus.

c) Les arbres de décision (decision trees)

Les arbres de décision sont la modélisation d'une classification. Ils apprennent à partir d'observations qu'on appelle des exemples. Un exemple est représenté par une série d'attributs et une classe associée, on doit connaître la classe parce que les arbres de décision travaillent sur la classification en mode supervisée [18]. Les arbres de décision sont un bon moyen d'illustrer le raisonnement pour distinguer les similitudes et les différences entre les attributs des exemples du jeu de données, ils sont souvent utilisés par les statisticiens pour illustrer le résultat d'une analyse.

Un arbre de décision est composé de nœuds en arborescence, le nœud à base de l'arbre est appelé la racine, chacun des nœuds sous la racine est soit une feuille ou un sous arbre.



**Figure 6 Schéma d'un arbre de décision [18]**

Dans la figure 6, les nœuds B, D et E sont des nœuds terminaux et le nœud C est un sous arbre du nœud A.

**Une feuille** est un nœud terminal qui représente le résultat d'une classification. La racine d'un sous arbre est étiquetée avec l'attribut qui a été choisi, les branches sont étiquetées avec les différentes valeurs que peut prendre l'attribut choisi pour le nœud.

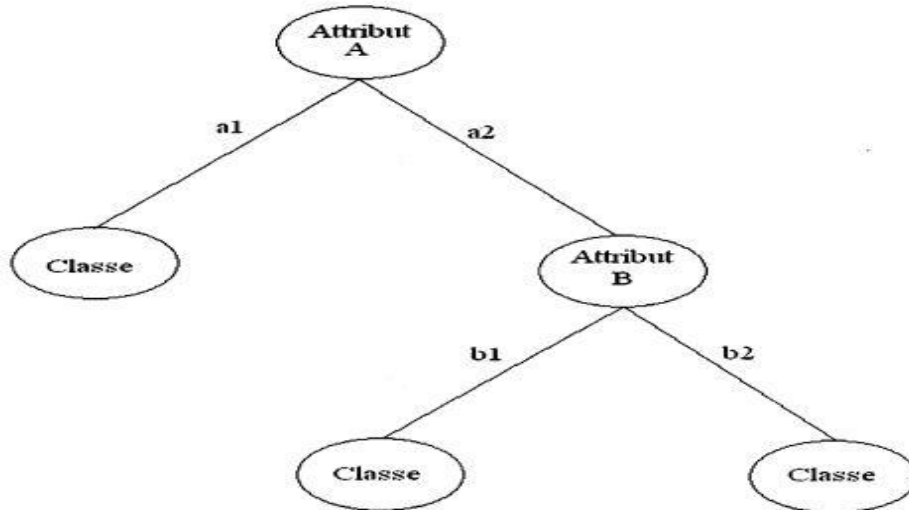


Figure 7: Schéma d'un arbre de décision pour les attributs et les classes [18].

Dans l'arbre de décision de la figure 7, les attributs A et B ont chacun deux valeurs distinctes, lorsque les exemples avec l'attribut A égal à a1, ils correspondent à une seule classe.

Dans le cas où l'attribut A est égal à a2, les exemples correspondent à deux classes différentes, on a besoin alors de prendre l'attribut B pour diviser les exemples dans leurs classes respectives.

Il existe plusieurs algorithmes de construction d'arbre, les plus populaires sont ID3, C4.5, CART et CHAID. Ces algorithmes permettent de créer des arbres de décision avec une taille la plus petite que possible, et ce, de façon à créer des règles de décision simples. Plus un arbre de décision est grand, plus les règles sont complexes. Les algorithmes de construction d'arbres choisissent les attributs toujours par rapport aux classes.

Le principal problème qui se pose lors de la construction d'un arbre de décision est de savoir comment choisir ou sélectionner le meilleur attribut pour le nœud racine et qui sépare mieux l'ensemble de données. Pour résoudre ce problème il existe une technique appelée Mesure de sélection d'attribut ou ASM qui contient deux mesures principales et populaires suivantes :

- **Indice de Gini** : c'est une mesure d'impureté ou de pureté utilisée lors de la création d'un arbre de décision dans l'algorithme CART (arbre de classification et régression). L'indice de Gini peut être calculé à l'aide de la formule suivante :

$$\text{Gini}(S) = 1 - \sum_{j=1}^n p_j^2$$

Équation 9: calcul de l'indice de Gini

Où S est l'ensemble d'échantillon, n le nombre de classe à prédire et Pj est la fréquence de la classe j dans l'ensemble d'échantillon.

- **Gain d'information** : C'est une mesure de sélection utilisé pour mesurer la pureté d'un attribut, et définir le degré de désorganisation dans un système appelé l'entropie. Si l'échantillon est complètement homogène, l'entropie est nulle et si l'échantillon est également divisé (50% - 50%), il a une entropie de un. L'entropie peut être calculée à l'aide de la formule suivant :

$$H(S) = -p.\log_2(p) - q.\log_2(q)$$

**Équation 10:calcul de l'entropie d'un attribut de l'ensemble.**

Où S est l'ensemble d'échantillon et p et q respectivement la probabilité de succès et d'échec dans l'ensemble d'échantillon.

Pour chaque caractéristique, on calcule le gain d'information.

$$IG(S, f_j) = H(S) - \sum_{S_{jk} \in S_j} P(S_{jk})H(S_{jk})$$

Où:

$$P(S_{jk}) = \frac{|S_{jk}|}{|S|}$$

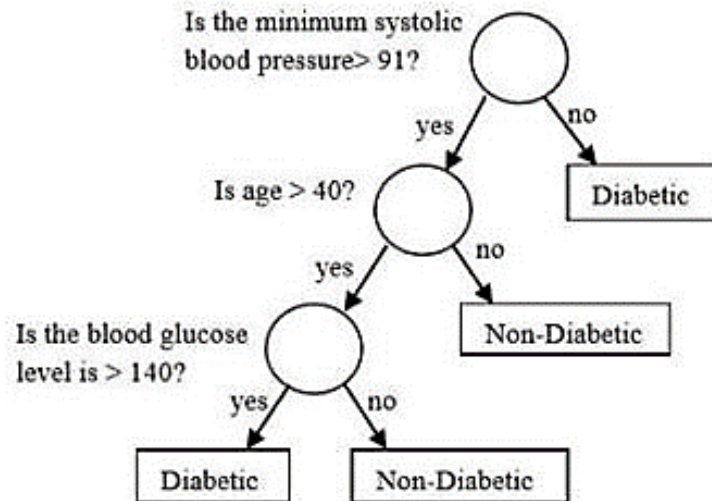
**Équation 11: Calcul du gain d'information**

**L'algorithme de construction d'arbre de décision se divise en 3 étapes :**

- **Sélectionner le meilleur attribut (nœud racine)** : pour chaque attribut le gain d'information est calculé et celui qui est maximal est sélectionné et des branches sont créées pour chaque valeur de cet attribut. La deuxième et la troisième étape se produisent lorsqu'on ne respecte pas les critères de la première.
- **Continuez la division** : pour chaque branche s'étendant à partir de nœud, en répétant récursivement le processus.
- **Arrêter la division si** :
  - On obtient un nœud pur, c'est à dire un nœud qui ne contient que des points de données positifs ou négatifs.
  - On obtient très peu de points dans un nœud.
  - On atteint une certaine profondeur de l'arbre. [19]

Illustrons tous les paragraphes ci-haut par un exemple pour une meilleure compréhension.

C'est un petit exemple pour prédire si une personne est diabétique ou non, ce modèle contient trois attributs qui sont << minimum systolic Blood pressure >>, << âge >> et << glucose >> avec deux classes << diabetic >> et << non-diabetic >>.



**Figure 8:Arbre de décision qui répond à la question si une personne est diabétique ou non. [19]**

- L'interprétation de l'exemple :

Les arbres de décision sont bien adaptés aux problèmes de catégorisation où les attributs sont vérifiés pour déterminer une catégorie finale à cause de sa construction naturelle <<si ... alors ... sinon ... >>. Lecture de l'exemple :

**Si** minimum systolic blood pressure > 91 = no, **alors** : personne = diabetic;

**Si** minimum systolic blood pressure > 91 = yes **AND** age > 40 = no, **alors**: personne = non-diabetic;

**Si** minimum systolic blood pressure > 91 = yes **AND** age > 40 = yes **AND** glucose =no, **alors**: personne = non-diabetic;

**Si** minimum systolic blood pressure > 91 = yes **AND** age > 40 = yes **AND** glucose =yes, **alors**: personne = diabetic;

#### d) Random Forest (forêts aléatoires)

Random Forest ou forêts aléatoires est un algorithme d'apprentissage supervisé très populaire. Il est également utilisé pour les problèmes de régression ou de classification. Basé sur un ensemble des algorithmes d'apprentissage, qui est un processus de combinaison de plusieurs algorithmes pour résoudre un problème complexe et améliorer les performances du modèle. C'est un algorithme qui crée de nombreux arbres de décision (c'est la raison pour laquelle

il est appelé une forêt) sur divers sous-ensembles de l'ensemble de données. Elle prend la prédiction de chaque arbre et sur la base des votes majoritaires des prédictions, et elle prédit le résultat final. [20]

La figure suivant explique le fonctionnement et la structure d'algorithme :

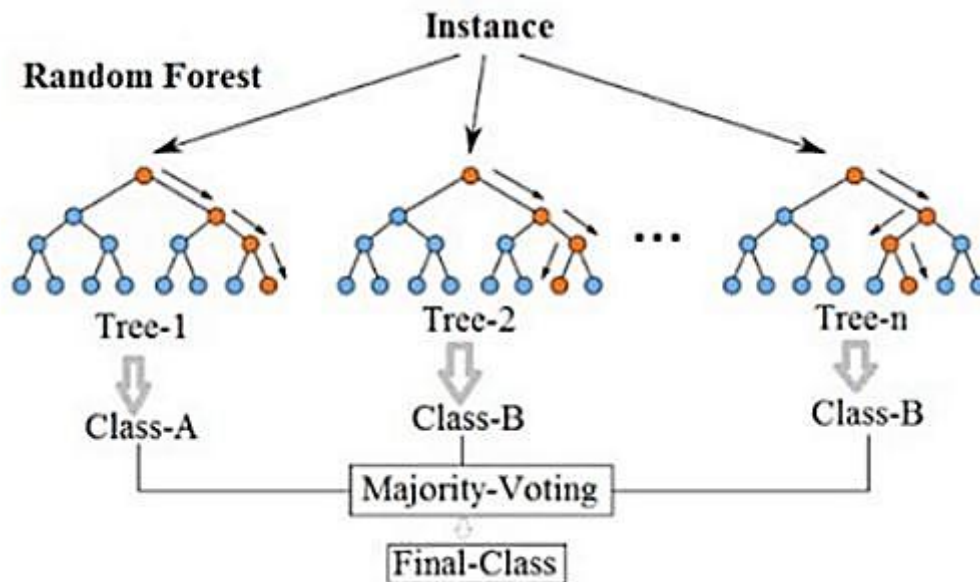


Figure 9: Structure de l'algorithme Radom Forest [20]

- **Algorithme de construction de Random Forest :**
  - Sélectionnez des échantillons aléatoires à partir d'un ensemble de données d'entraînement.
  - Créer des arbres de décision pour chaque échantillon (sous-ensembles). Ensuite on obtient le résultat de prédiction de chaque arbre de décision
  - Pour les nouveaux points le vote sera effectué pour chaque résultat prédit.
  - Sélectionnez le résultat de prédiction le plus voté comme résultat de prédiction final. [21]
- **Avantage de Random Forest :**
  - Il s'agit de l'un des algorithmes d'apprentissage les plus précis disponibles. Pour de nombreux ensembles de données, il produit un classificateur très précis.
  - Il fonctionne efficacement sur de grandes bases de données.
  - Il dispose d'une méthode efficace pour estimer les données manquantes et maintient la précision lorsqu'une grande partie des données sont manquantes.

- **Inconvénient de Random Forest :**

Le principal inconvénient de l'algorithme Random Forest est qu'un grand nombre d'arbres peut rendre l'algorithme trop lent et inefficace pour les prédictions en temps réel. En général, ces algorithmes sont rapides à entraîner, mais assez lents à créer des prédictions une fois qu'ils sont formés. Une prévision plus précise nécessite plus d'arbres, ce qui entraîne un modèle plus lent.

#### 2.4.2. Le clustering.

La méthode de clustering est une technique d'apprentissage automatique non supervisée qui permet de regrouper des données similaires en fonction de leurs caractéristiques. Elle est utilisée dans de nombreux domaines, tels que la science des données, le marketing, la finance et la médecine. [22]

Le clustering consiste à trouver un ensemble de groupes, appelés clusters, tels que les données dans chaque cluster sont plus similaires les unes aux autres qu'aux données des autres clusters. La similarité est mesurée en fonction d'une distance ou d'une similarité.

Il existe de nombreuses méthodes de clustering différentes. Les méthodes les plus courantes sont les suivantes :

- **Les méthodes hiérarchiques** divisent l'ensemble de données en clusters de plus en plus petits, en utilisant une mesure de similarité pour regrouper les données les plus proches.
- **Les méthodes non hiérarchiques** regroupent les données en clusters de manière simultanée, en utilisant une fonction d'optimisation pour trouver la configuration qui minimise la distance entre les données du même cluster.

Le choix de la méthode de clustering appropriée dépend de plusieurs facteurs, tels que la nature des données, le nombre de clusters souhaité et la mesure de similarité utilisée.

Voici quelques exemples d'utilisation de la méthode de clustering :

- **En marketing**, le clustering peut être utilisé pour identifier des groupes de clients similaires en fonction de leurs habitudes d'achat. Cela peut aider les entreprises à cibler leurs campagnes marketing de manière plus efficace.
- **En finance**, le clustering peut être utilisé pour identifier des groupes de valeurs similaires en fonction de leurs performances. Cela peut aider les investisseurs à diversifier leurs portefeuilles de manière plus efficace.

- **En médecine**, le clustering peut être utilisé pour identifier des groupes de patients similaires en fonction de leurs symptômes. Cela peut aider les médecins à identifier des traitements plus efficaces.

### 2.4.3. La régression.

La régression est une technique statistique qui permet de modéliser la relation entre une variable dépendante et une ou plusieurs variables explicatives. La variable dépendante est la variable que l'on souhaite prédire, tandis que les variables explicatives sont les variables qui influencent la variable dépendante. Elle est souvent utilisée dans la recherche médicale pour identifier les facteurs de risque du diabète et pour développer des outils de dépistage précoce.

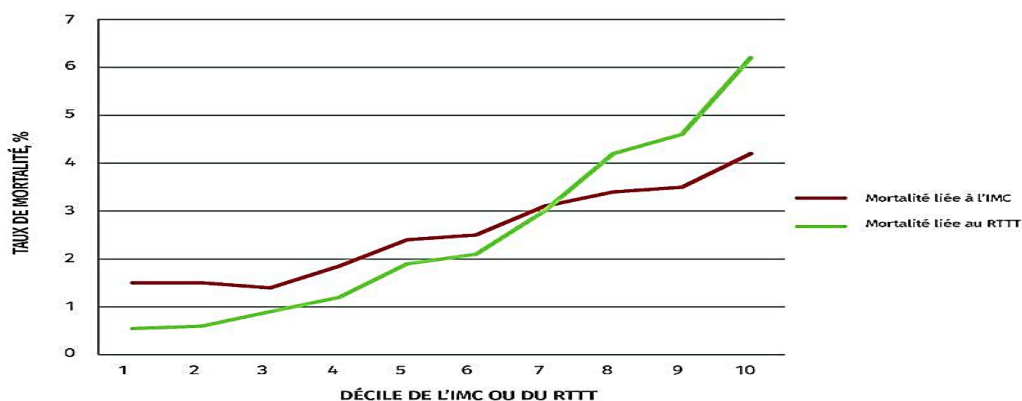
Il existe de nombreuses techniques de régression différentes, chacune avec ses propres avantages et inconvénients. Les types de régression les plus courants sont les suivants :

#### 2.4.3.1. La régression linéaire

Dans la prédiction d'une maladie, la régression linéaire est utilisée pour modéliser la relation entre des facteurs de risque connus et le risque de développer la maladie. Par exemple, on peut utiliser la régression linéaire pour prédire le risque de diabète en fonction de facteurs tels que l'âge, le sexe, l'IMC, l'historique familial de diabète et d'autres facteurs.

La régression linéaire suppose que la relation entre la variable dépendante (le risque de maladie) et les variables explicatives (les facteurs de risque) est linéaire. Autrement dit, la valeur du risque de maladie peut être représentée par une droite ou une courbe linéaire en fonction des valeurs des facteurs de risque.

La figure 10 ci-dessous illustre la régression linéaire dans la prédiction du diabète :



**Figure 10: Courbe linéaire représentant la relation entre le risque de diabète et le taux de mortalité [23]**

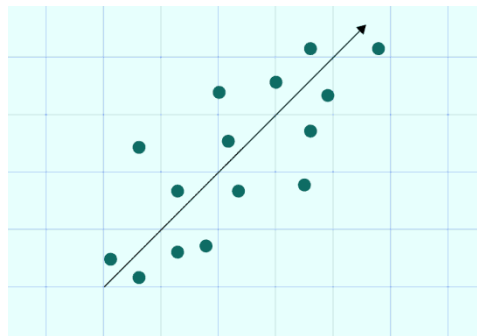
L'axe des abscisses représente l'âge et l'axe des ordonnées représente le risque de diabète. La courbe linéaire représente la relation entre ces deux variables.

#### 2.4.3.2. La régression non linéaire

La régression non linéaire est une technique d'apprentissage automatique qui permet de modéliser la relation entre une variable dépendante et une ou plusieurs variables explicatives, lorsque cette relation n'est pas linéaire. [24]

En d'autres termes, la régression non linéaire permet de modéliser des relations complexes entre les variables, qui ne peuvent pas être représentées par une simple droite ou une courbe linéaire.

La figure 11 ci-dessous fait une représentation des données d'un dataset du diabète en régression non linéaire



**Figure 11: Régression non-linéaire des données [24]**

Les points de l'image représentent les données du dataset Diabète Pred. Chaque point représente une observation, c'est-à-dire une personne. L'axe horizontal de l'image représentant l'âge, et l'axe vertical représente le risque de diabète.

#### 2.4.3.3. La régression polynomiale

Dans l'apprentissage automatique, la régression polynomiale est utilisée pour modéliser des relations non linéaires entre des variables. Elle est souvent utilisée dans des applications de classification et de prédiction. [24]

Par exemple, on peut utiliser la régression polynomiale pour prédire le risque de développer une maladie, le prix d'un produit ou la probabilité qu'un client effectue un achat.

La régression polynomiale est une technique puissante qui peut être utilisée pour améliorer la précision des modèles d'apprentissage automatique.

Voici une image illustrative d'une courbe polynomiale du second degré :

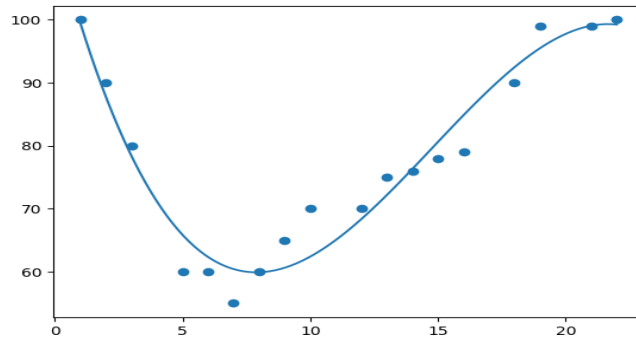


Figure 12: Courbe polynomiale du second degré [24]

La courbe polynomiale du second degré représente la relation entre les deux variables x et y.

#### 2.4.3.4. La régression logistique

La régression logistique est un modèle statistique qui est utilisé pour prédire la probabilité d'un événement binaire. Un événement binaire est un événement qui peut avoir deux valeurs possibles, par exemple "oui" ou "non", "vrai" ou "faux", "malade" ou "en bonne santé". Dans la régression logistique, ce n'est pas la réponse binaire (malade/ en bonne santé) qui est directement modélisée, mais la probabilité de réalisation d'une des deux modalités (être malade par exemple). Cette probabilité de réalisation ne peut pas être modélisée par une droite car celle-ci conduirait à des valeurs inférieures à zéro ou supérieures à un. Ce qui est impossible puisqu'une probabilité est forcément bornée par 0 et 1. [25] Cette **probabilité**, est alors modélisée par une **courbe sigmoïde**, bornée par 0, et 1 comme la montre la figure ci-dessous.

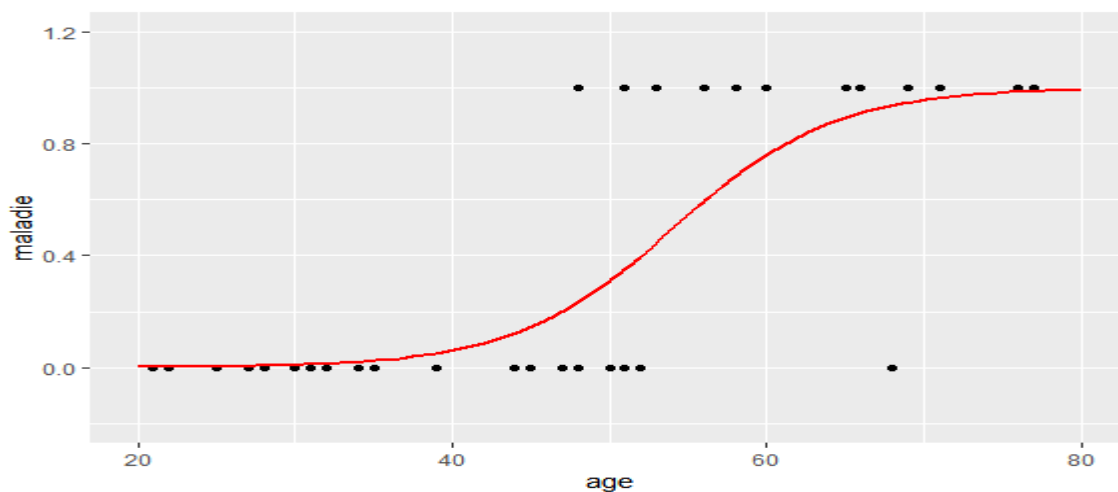


Figure 13: courbe sigmoïde de régression linéaire [25]

Cette courbe sigmoïde est définie par la **fonction logistique**, d'équation :

$$f(x) = \frac{\exp(x)}{1 + \exp(x)}$$

**Équation 12: fonction logistique [25]**

## 2.5. Application des méthodes d'Apprentissage automatique dans la Prédiction du Diabète.

Comme nous l'avons vu précédemment, l'apprentissage automatique est utilisé dans la médecine pour augmenter la précision, le temps de réponse et beaucoup d'autres avantages.

### 2.5.1. La justesse

La justesse (Accuracy) est le taux de réussite global de l'algorithme défini par l'équation suivante :

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

**Équation 13: calcul de l'accuracy**

**TP** : True Positive. **TN** : True Negative. **FN** : False négatif. **FP** : False Positive.

Dans notre cas d'étude, la signification de TP, TN, FP et FN est comme suit :

- TP : signifie qu'une personne est réellement diabétique et elle a été prédite qu'elle est diabétique.
- TN : signifie qu'une personne est réellement non diabétique et elle a été prédite qu'elle est non diabétique.
- FP : signifie qu'une personne est réellement non diabétique et elle a été prédite qu'elle est diabétique.
- FN : signifie qu'une personne est réellement diabétique et elle a été prédite qu'elle est non diabétique.

### 2.5.2. Etude sur la précision de prédiction du diabète.

Une étude intitulée "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare" réalisée par **Muhammad Azeem Sarwar** et **Nasir Kamal** en 2018 a été faite sur les méthodes appropriées pour prédire l'état du diabète des patients. Cette étude a conclu que SVM et KNN sont appropriés pour prédire l'état du diabète des patients entre six algorithmes d'apprentissage automatique. Ces algorithmes sont : K-Nearest Neighbors (KNN), Support

Vector Classifier(SVM), Logistic Regression(LR), Decision Tree Classifier(DT), Gaussian Naive Bayes(NB) et Random Forest (RF). Tous ces algorithmes ont été appliqués au Dataset PIMA Indian comprenant 768 enregistrements et 9 attributs. Dans ce travail, le principal paramètre d'évaluation entre les algorithmes est la justesse de prédiction mais la précision doit être trop importante de prédiction. [26]

**Tableau 0-1:Les résultats de la précision des algorithmes d'étude [26].**

Algorithmes	KNN	SVM	NB	LR	DT	RF
Précision	77%	77%	74%	74%	71%	71%

Afin d'évaluer la précision de ces méthodes, deux méthodes d'évaluation sont applicables :

- Méthode Train/ Test Split

Ce modèle est celui utilisé dans ce travail, elle consiste à fractionner l'ensemble de données en deux parties :

- **Partie d'entraînement** : pour former le modèle.
- **Partie de test** : pour tester le modèle et évaluer la précision.

$$Précision = \frac{VP}{VP + FP}$$

#### Équation 14: Mesure de la précision du modèle

- Méthode de Validation croisée.

La validation croisée consiste à diviser le jeu de données d'entraînement en plusieurs sous-ensembles de taille égale ou inégale. Le modèle est entraîné sur un sous-ensemble à la fois et ses performances sont évaluées sur les autres sous-ensembles. [24]

Il existe de nombreuses méthodes de validation croisée, chacune avec ses propres avantages et inconvénients. Les méthodes les plus courantes sont les suivantes [24] :

- **K-fold cross-validation** : le jeu de données est divisé en K sous-ensembles. Le modèle est entraîné sur K-1 sous-ensembles et ses performances sont évaluées sur le Kème sous-ensemble.
- **Leave-one-out cross-validation** : le jeu de données est divisé en K sous-ensembles, chacun contenant un seul point de données. Le modèle est entraîné sur les K-1 sous-ensembles et ses performances sont évaluées sur le Kème sous-ensemble.

- **Stratified cross-validation** : le jeu de données est divisé en K sous-ensembles, chacun ayant la même distribution que le jeu de données d'origine. Cette méthode est utile lorsque le jeu de données d'origine contient des classes déséquilibrées.

Le choix de la méthode de validation croisée dépend des caractéristiques du jeu de données et des objectifs de l'évaluation.

### 2.5.3. Le choix de la méthode pour notre travail

Nous avons fait un entraînement du modèle comme nous le verrons dans le chapitre suivant et sur 3 méthodes la régression logistique est celle qui nous a convaincu avec une précision de prédiction du diabète convenable et une haute justesse ou accuracy. Hormis la nature même des données contenues dans le dataset et les expériences qu'ont montrées les scientifiques sur la prédiction du diabète, la méthode de régression logistique montre une grande justesse avec la méthode test split comparativement aux autres comme le montre le tableau 2-2 . En médecine, il est souvent préférable de privilégier la justesse pour éviter de rater des patients malades, mais cela dépendra de notre contexte spécifique (p. ex., coût des faux positifs ou des faux négatifs)

- Si la **priorité est de diagnostiquer tous les cas possibles**, ce qui est le cas de notre travail, favorisez la **justesse** et ajustez le seuil pour maximiser la sensibilité.
- Si la **priorité est d'éviter les faux diagnostics**, favorisez la **précision** pour minimiser les faux positifs.

**tableau 0-2:comparaison de la précision et de la justesse des différentes méthodes**

<b>Méthode</b>	<b>Precision.</b>	<b>Accuracy.</b>
XGBoost simple	75.65	97.14
XGBoost avec early-stopping = 5	75.40	97.25
Random Forest	71.30	97.04
Logistic Regression	74.24	97.60

### D'autres mesures de performance adaptées pour évaluer notre modèle :

- **Sensibilité (Recall)** : Mesure la capacité du modèle à détecter les vrais cas de diabète (minimiser les faux négatifs).

$$Recall = \frac{VP}{VP + FN}$$

#### Équation 15: mesure du rappel ou la sensibilité

- **F1-score** : Combine la sensibilité et la précision pour un compromis équilibré.

$$F1 - score = \frac{2 \cdot (\text{precision} \cdot \text{recall})}{(\text{precision} + \text{recall})}$$

#### Équation 16 : calcul de F1-score

- **AUC-ROC** : Permet d'évaluer les performances globales du modèle en termes de capacité à discriminer entre malades et non-malades.

## 2.6. Conclusion partielle

Dans ce chapitre, nous avons présenté les concepts fondamentaux de l'apprentissage automatique appliqué au diagnostic du diabète. Ces informations ont orienté notre sélection du modèle d'intelligence artificielle qui sera utilisé dans la poursuite de notre étude. Il s'agit de la régression logistique. De plus, elles nous seront essentielles dans la conception de notre outil d'assistance au diagnostic du diabète et de suivi des patients utilisateurs. Ce chapitre sert de pont vers la conception et la réalisation de l'application web qui feront l'objet du chapitre trois.

# Chapitre 3 : Conception et réalisation de l'application web.

## 3.1. Introduction

Dans ce chapitre, nous présentons les modèles et diagrammes pour une bonne compréhension du fonctionnement de notre outil d'aide au diagnostic du diabète en utilisant l'apprentissage automatique. En ce qui concerne les modèles, nous avons sélectionné le modèle d'entraînement pour notre Intelligence Artificielle dans le chapitre précédent, optant finalement pour la méthode de régression logistique. Ce chapitre abordera les aspects fonctionnels du système, la modélisation de la base de données, et se penchera ensuite sur la mise en œuvre de l'application web.

## 3.2. Présentation du système

### 3.2.1. Cadre d'outil

L'outil d'aide au diagnostic du diabète (que nous avons appelé Santé Plus) est un système qui permet de faire un diagnostic rapide grâce à l'apprentissage automatique ; Sa contribution au diagnostic sera précieuse pour le médecin, facilitant ainsi le suivi de la maladie du patient diabétique lors de la remise de ses échantillons.

### 3.2.2. Cahier de charge

Il est indispensable de définir un cahier de charge qui va servir de guide dans la conception de notre système. C'est en quelque sorte la volonté de l'utilisateur qui veut avoir cet outil. Soit la volonté du médecin en chef, du directeur de l'entreprise ou d'une toute autre personne qui veut en avoir.

L'application à implémenter dans ce travail doit permettre de :

- Faire un diagnostic rapide du diabète ; pour cela elle doit :
  - Avoir un modèle d'intelligence artificiel prêt à utiliser pour le diagnostic.
  - Avoir la possibilité d'insérer les valeurs des examens et les soumettre.
  - Afficher les résultats aux sorties de modèles.
  - Enregistrer le patient et son état.
- Et faire le suivi de patients ; pour s'y prendre elle doit être capable de (d') :
  - Enregistrer les utilisateurs et les médecins et le patients;

- Authentifier et d'identifier tous les utilisateurs de façon sécurisée ;
- Enregistrer les données des consultations et de test ;
- Afficher les données de consultations et de test ;
- Permettre au médecin de faire un suivi de santé du patient.

### 3.2.3. Architecture

L'application qui constitue notre travail prend en considération deux facteurs :

- 1) Un modèle de l'apprentissage automatique
- 2) L'enregistrement des données qui vont servir pour faire le suivi et voir l'historique.

Ce dernier demande une base de données et un API (Application Programming Interface) qui est l'interface homme-machine qui permet la communication entre les utilisateurs et le système modèle et base des données. Cet API sera fait en web et la base de données qui sera fait en SQLite (un type de base de données présent dans Django). La figure ci-dessous illustre cette architecture.

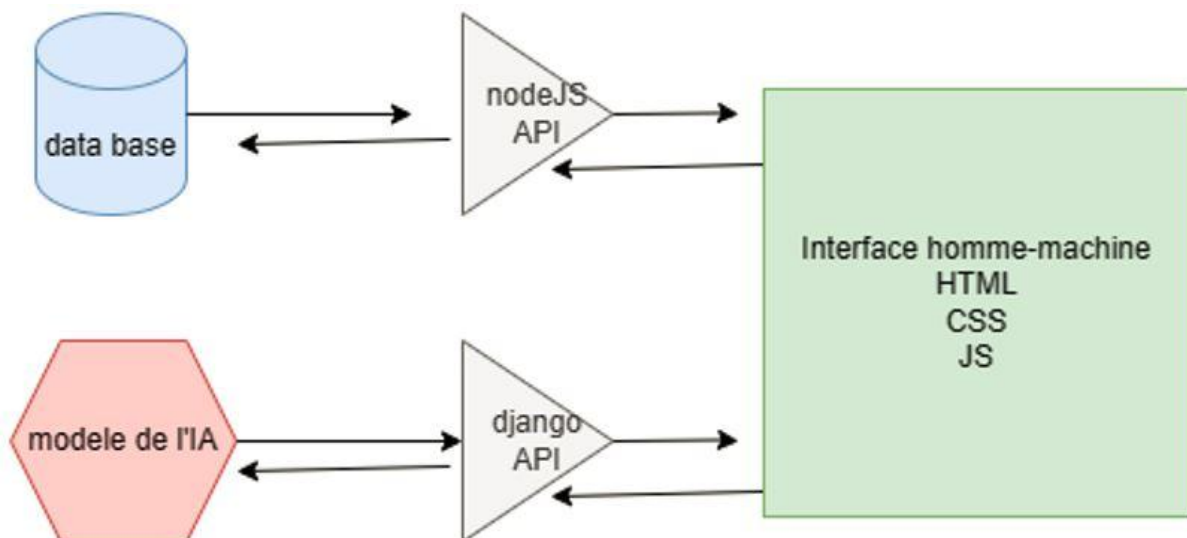


Figure 14: Architecture de l'outil SantePlus.

### 3.3. Le modèle de l'intelligence artificiel

Les données d'entraînement sont issues d'une plateforme de données du nom de Kaggle.

Notre DataSet est un fichier de type csv qui contient cent milles donnés.

### 3.3.1. Les entrées et les sorties

Les entrées pour notre modèle seront des chiffres ; et comme sorties des chiffres binaires 1 pour le cas où le résultat de diagnostic est « positif » et 0 dans le cas contraire.

Huit des colonnes du dataset dont « gender », « hypertension », « heart disease », « smoking history », « bmi », « HbA1c\_level » et « blood\_glucose\_level » sont pris comme des entrées que nous devons récupérer par un formulaire de notre application et la dernière colonne, « diabete » est la sortie.

Les colonnes qui ont des valeurs qui ne sont pas des nombres comme « gender » et « smoking history » doivent être encodé avant de faire l'entraînement sur ce dataset. LabelEncoder de la librairie sklearn est l'encodeur qui a été utilisé dans ce travail mais il en existe beaucoup.

### 3.3.2. Visualisation du dataset

Le tableau ci-dessous montre cinq des lignes de ce dataset.

**Tableau 0-1 Affichage de cinq des données du dataset avant encodage**

	Gender	Age	Hypertension	heart_disease	smoking_history	Bmi	HbA1c_level	blood_glucose_level	Diabete
0	Female	80.0	0	1	Never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	Never	27.32	5.7	158	0
3	Female	36.0	0	0	Current	23.45	5.0	155	0
4	Male	76.0	1	1	Current	20.14	4.8	155	0

Après l'utilisation de l'encodeur, les deux colonnes « gender » et « smoking history » ont des nombres comme valeurs et l'entraînement devient aisé.

En considérant les types de données d'entraînement présents dans notre ensemble de données, ainsi que les entrées et sorties attendues, et en tenant compte de la précision, nous optons pour l'application du modèle d'apprentissage automatique basé sur la régression logistique.

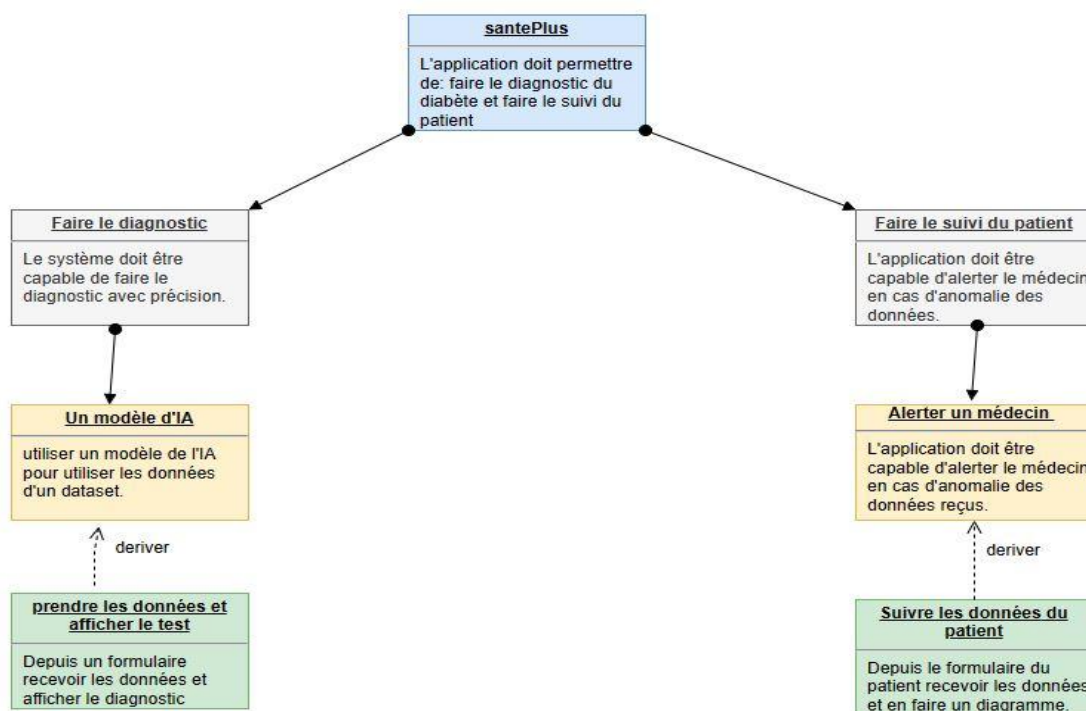
### 3.4. Conception fonctionnelle du système

Dans cette partie de ce travail, nous présentons la conception du système. Celle-ci permet de représenter et de stimuler son fonctionnement pour passer à la programmation de l'application. Dans le cadre de ce travail, nous allons utiliser les diagrammes suivants : le diagramme d'exigence, le diagramme de cas d'utilisation ainsi que les diagrammes d'activités. Nous avons utilisé le logiciel Draw.io [15] [16] [17] pour dessiner ces diagrammes.

#### 3.4.1. Diagramme d'exigence

Les exigences sont dites spécifications fonctionnelles et permettent de spécifier une capacité ou une contrainte qui doit être satisfaite par une application ou un système donné. Le diagramme d'exigence permet de représenter graphiquement les spécifications fonctionnelles dans le modèle. Une spécification fonctionnelle peut spécifier une fonction que le système devra réaliser ou une condition de performance, de fiabilité, de sécurité, etc.

Pour le cas de notre système, L'exigence principale portera le nom du système : « santePlus » et les autres exigences du second ordre (enregistrer ; authentifier et d'identifier un utilisateur; enregistrer les données des consultations et de test ; afficher les données de consultations et de test ; voir historique de santé d'un patient.) sont liées à santePlus par les



relations de contenance.

**Figure 15: Diagramme d'exigences de l'application SantePlus.**

### 3.4.2. Diagramme des cas d'utilisation

Les diagrammes de cas d'utilisation sont des diagrammes utilisés pour donner une vision globale du comportement fonctionnel d'un système logiciel. Ils sont utiles pour des présentations auprès de la direction ou des acteurs d'un projet, mais pour le développement, les cas d'utilisation ne sont pas appropriés.

Un cas d'utilisation (use case, ou UC) représente un ensemble de séquences d'actions qui sont réalisées par le système et qui produisent un résultat observable intéressant pour un acteur particulier. Il permet de décrire ce que le futur système devra faire, sans spécifier comment il le fera. Chaque cas d'utilisation doit être relié à au moins un acteur (une entité externe qui interagit avec le système, comme une personne humaine ou un robot). Un acteur doit toujours être décrit par son rôle dans un système, ce rôle décrit les besoins et les capacités de l'acteur.

## Diagramme

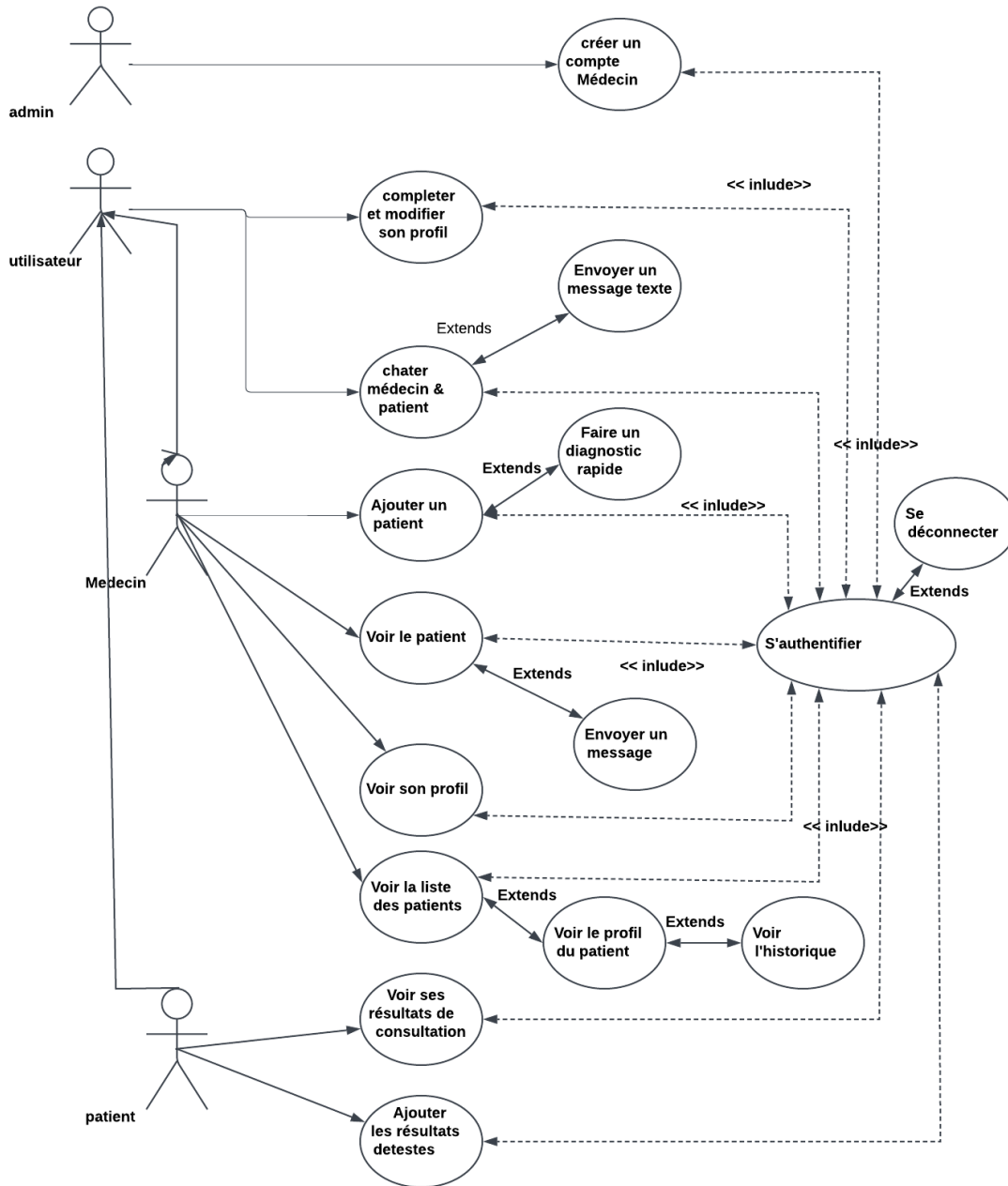


Figure 16: Diagramme des cas d'utilisation de SantePlus.

### 3.4.3. Documentation des cas d'utilisation

De la figure du diagramme des cas d'utilisation, si nous prenons un cas d'utilisation après l'autre pour faire la documentation nous avons les tableaux suivants :

#### ➤ Authentification

**Tableau 0-2:Cas d'utilisation 001 - authentification**

<b>Nom du cas d'utilisation</b>	S'authentifier
<b>Identifiant</b>	001
<b>Description</b>	Permet la connexion au système pour tous utilisateurs
<b>Acteur</b>	Utilisateur/admin
<b>Pré condition</b>	Aucun
<b>Enchaînement principale</b>	<ul style="list-style-type: none"><li>○ L'utilisateur clique sur le bouton connexion</li><li>○ Il remplit le formulaire de connexion au système</li><li>○ Il clique sur se connecter</li></ul>
<b>Post condition</b>	Un utilisateur est connecté
<b>Enchaînement secondaire</b>	Si le formulaire est mal rempli, la connexion échoue

#### ➤ Déconnexion

**Tableau 0-3:Cas d'utilisation 002 – déconnexion**

<b>Nom du cas d'utilisation</b>	Se déconnecter
<b>Identifiant</b>	002
<b>Description</b>	Permet l'utilisateur de se déconnecter au système
<b>Acteur</b>	Utilisateur/admin
<b>Pré condition</b>	Un utilisateur est connecté
<b>Enchaînement principale</b>	L'utilisateur clique sur le bouton se déconnecter
<b>Post condition</b>	Un utilisateur est déconnecté
<b>Enchaînement secondaire</b>	Aucu

➤ **Création compte médecin.**

**Tableau 0-4:Cas d'utilisation 003 – Créer un compte Médecin**

<b>Nom du cas d'utilisation</b>	Créer un compte médecin
<b>Identifiant</b>	003
<b>Description</b>	Permet de créer un compte médecin
<b>Acteur</b>	Administrateur
<b>Pré condition</b>	Un utilisateur est connecté comme administrateur
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ L'administrateur clique Ajouter un médecin</li> <li>○ Il complète le formulaire de création d'un docteur</li> <li>○ Il clique sur Ajouter</li> </ul>
<b>Post condition</b>	Un nouvel utilisateur est créé avec comme rôle médecin
<b>Enchaînement secondaire</b>	Si la création est mal faite, elle échoue

➤ **Création compte patient**

**Tableau 0-5:Cas d'utilisation 004 – Création compte patient**

<b>Nom du cas d'utilisation</b>	Ajouter un patient
<b>Identifiant</b>	004
<b>Description</b>	Permet de créer un compte patient
<b>Acteur</b>	Médecin
<b>Pré condition</b>	Un utilisateur est connecté comme Docteur
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ Le docteur clique Ajouter un patient</li> <li>○ Il complète le formulaire de création d'un patient</li> <li>○ Il clique sur Ajouter</li> </ul>
<b>Post condition</b>	Un nouvel utilisateur est créé avec comme rôle patient
<b>Enchaînement secondaire</b>	Si la création est mal faite, elle échoue

➤ **Profil**

**Tableau 0-6: Cas d'utilisation 005 – compléter et modifier son profil**

<b>Nom du cas d'utilisation</b>	<b>compléter et modifier son profil</b>
<b>Identifiant</b>	005
<b>Description</b>	Permet l'utilisateur de mettre à jour son profil
<b>Acteur</b>	Utilisateur
<b>Pré condition</b>	Un utilisateur est connecté comme patient/médecin
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ Le patient/médecin clique sur voir son profil</li> <li>○ Il complète le formulaire de son profil</li> <li>○ Il clique sur Ajouter</li> </ul>
<b>Post condition</b>	Profil du patient/médecin est mis à jour
<b>Enchaînement secondaire</b>	Aucun

➤ **Rechercher un patient**

**Tableau 0-7: Cas d'utilisation 006 – Voir un patient**

<b>Nom du cas d'utilisation</b>	Voir un patient
<b>Identifiant</b>	006
<b>Description</b>	Permet de faire la recherche d'un patient
<b>Acteur</b>	Médecin
<b>Pré condition</b>	<ul style="list-style-type: none"> <li>○ Un utilisateur est connecté comme médecin</li> <li>○ Le médecin doit être dans l'interface Liste des patients</li> </ul>
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ Le médecin clique sur l'icône de recherche.</li> <li>○ Il saisit les lettres de l'identité du patient</li> <li>○ Il clique sur recherche</li> </ul>
<b>Post condition</b>	La liste de patient est filtrée selon les textes entrés
<b>Enchaînement secondaire</b>	Si l'identité du patient n'existe pas, la page sera vide

➤ **Voir le patient**

**Tableau 0-8: Cas d'utilisation 007 – Voir la liste des patients**

<b>Nom du cas d'utilisation</b>	Voir la liste des patients
<b>Identifiant</b>	007
<b>Description</b>	Permet voir la liste des patients
<b>Acteur</b>	Médecin
<b>Pré condition</b>	Un utilisateur est connecté comme médecin
<b>Enchaînement principale</b>	Le médecin clique sur patients
<b>Post condition</b>	La liste des patients apparaît
<b>Enchaînement secondaire</b>	Aucun

➤ **Faire un diagnostic rapide**

**Tableau 0-9: Cas d'utilisation 008 – Faire un diagnostic rapide**

<b>Nom du cas d'utilisation</b>	Faire un diagnostic
<b>Identifiant</b>	008
<b>Description</b>	Permet de faire un diagnostic pour un patient
<b>Acteur</b>	Médecin
<b>Pré condition</b>	Un utilisateur est connecté comme médecin
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ Le médecin clique sur diagnostic</li> <li>○ Il remplit le formulaire des examens faits</li> <li>○ Clique sur diagnostic</li> </ul>
<b>Post condition</b>	Le résultat de diagnostic s'affiche
<b>Enchaînement secondaire</b>	Si le formulaire de diagnostic est mal complété il y a erreur

➤ **Voir les résultats de diagnostic**

**Tableau 0-10: Cas d'utilisation 009 – Voir les résultats de diagnostic**

<b>Nom du cas d'utilisation</b>	Voir les résultats de diagnostic
<b>Identifiant</b>	009
<b>Description</b>	Permet de voir le profil et l'historique d'un patient
<b>Acteur</b>	Patient
<b>Pré condition</b>	Un utilisateur est connecté comme patient
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ Le patient ouvre son compte.</li> <li>○ Il clique sur l'onglet Résultats</li> </ul>
<b>Post condition</b>	Les résultats de consultations s'affichent
<b>Enchaînement secondaire</b>	Aucun

○ **Message**

**Tableau 0-11: Cas d'utilisation 010 –Envoyer un Message.**

<b>Nom du cas d'utilisation</b>	Envoyer un message
<b>Identifiant</b>	010
<b>Description</b>	Permet à l'utilisateur d'envoyer un texte à un autre utilisateur.
<b>Acteur</b>	Utilisateur
<b>Pré condition</b>	Un utilisateur est connecté
<b>Enchaînement principale</b>	<ul style="list-style-type: none"> <li>○ L'utilisateur clique sur Message</li> <li>○ Il complète le message à envoyer</li> <li>○ Il clique sur envoyer</li> </ul>
<b>Post condition</b>	Un nouveau message est envoyé à un utilisateur
<b>Enchaînement secondaire</b>	Aucun

### 3.4.4. Diagrammes d'activités de l'application SantePlus.

Un diagramme d'activité est un diagramme UML qui décrit la séquence d'actions d'un processus. Il est utilisé pour modéliser le comportement d'un système, en particulier les flux de contrôle et de données. [27]

Les diagrammes d'activité sont similaires aux organigrammes de traitement de l'information, mais ils offrent des fonctionnalités supplémentaires qui les rendent plus puissants. Par exemple, les diagrammes d'activité peuvent être utilisés pour représenter des actions conditionnelles, des itérations et des choix.

Pour le cas de notre travail, nous allons prendre quelques cas d'utilisations et en faire des diagrammes d'activité. Nos diagrammes ont leurs composants encadrés en trois colonnes (les actions de l'acteur, les opérations qui se passe au niveau de la base de données et la dernière l'interface).

Diagramme d'activité pour le Cas d'utilisation 001 - authentification

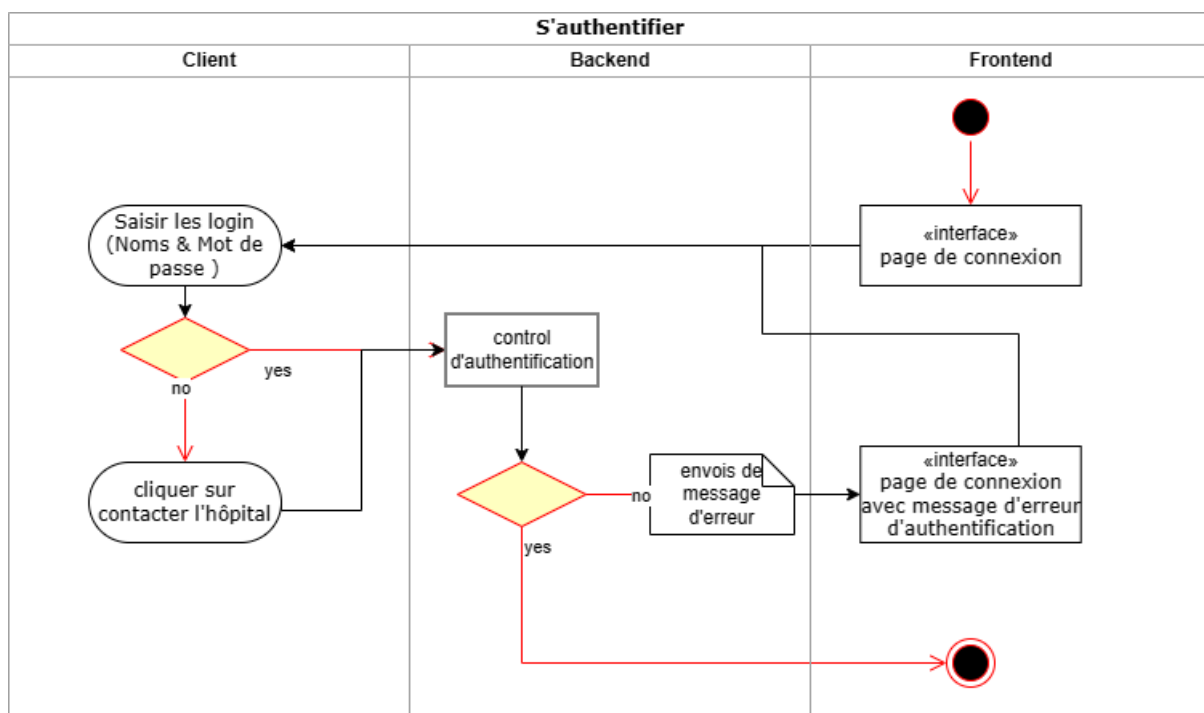


Figure 17: Diagramme d'activité pour s'authentifier dans SantePlus

Diagramme d'activité pour Cas d'utilisation 003 – Création compte docteur

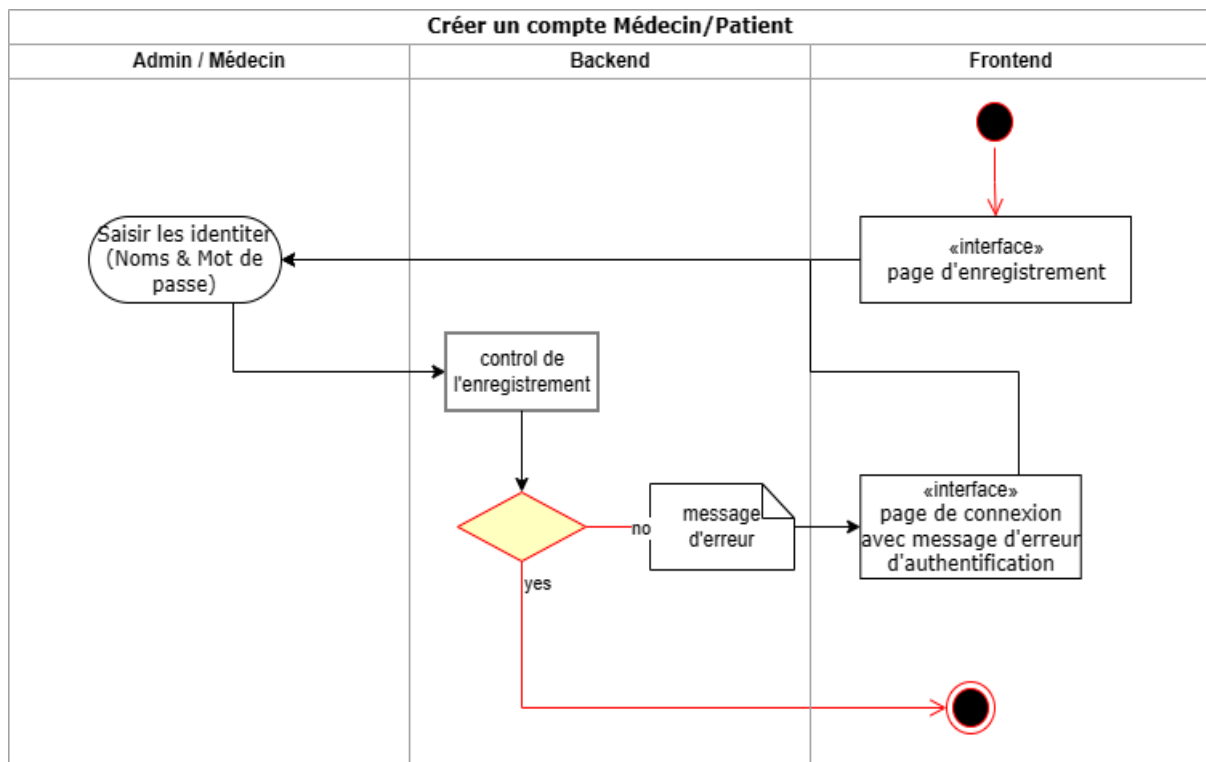


Figure 18:b) Diagramme d'activité pour créer un compte dans SantePlus

Diagramme d'activité pour cas d'utilisation 006 – Rechercher un patient

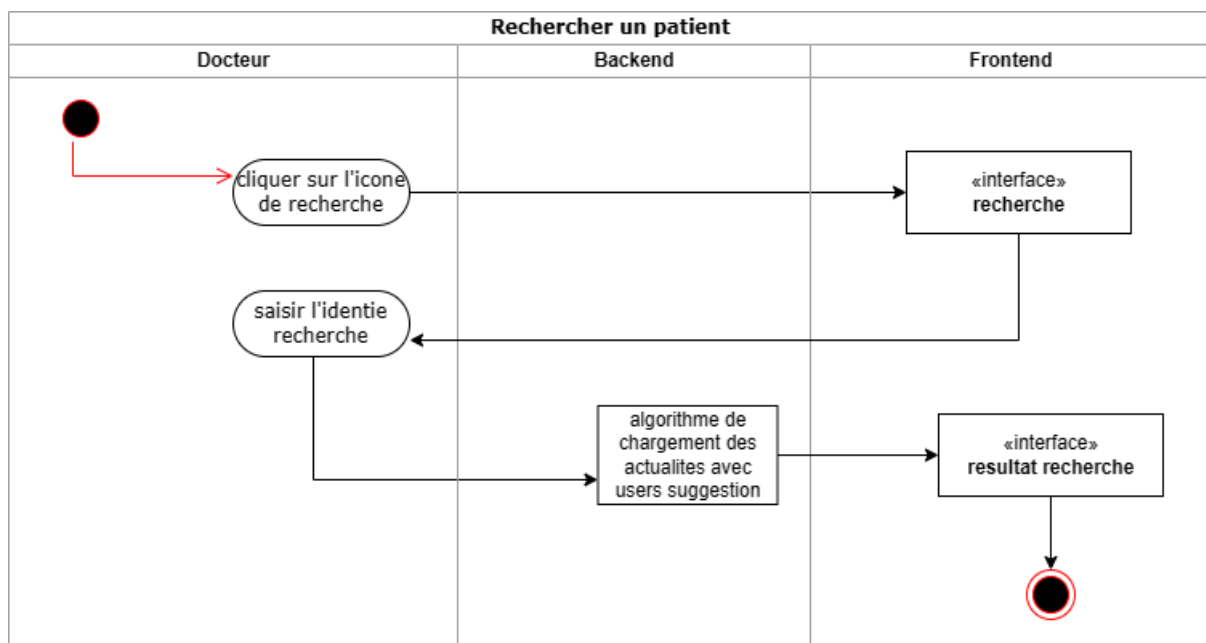


Figure 19 Diagramme d'activité pour rechercher un patient de SantePlus

Diagramme d'activité pour le Cas d'utilisation 005 – Profil

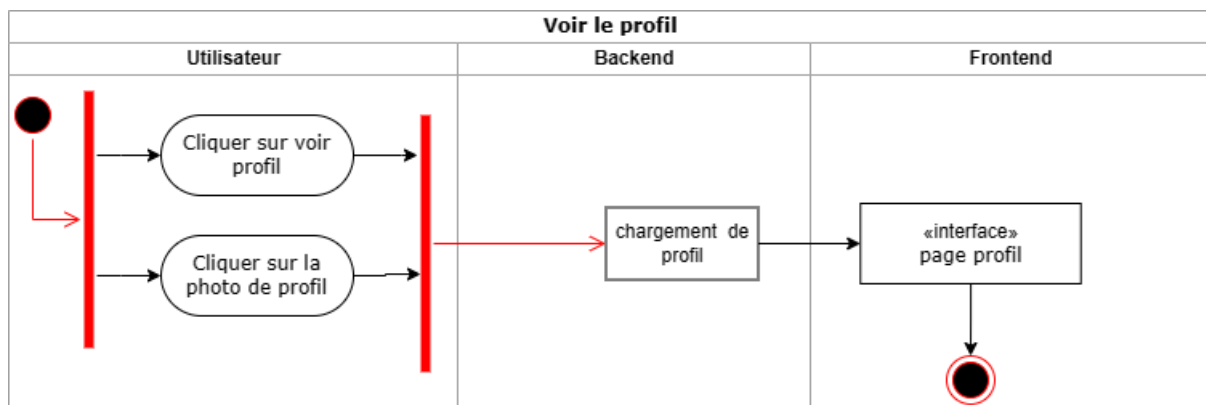


Figure 20: Diagramme d'activité pour voir le profil

Diagramme d'activité pour le cas d'utilisation 010 – Message.

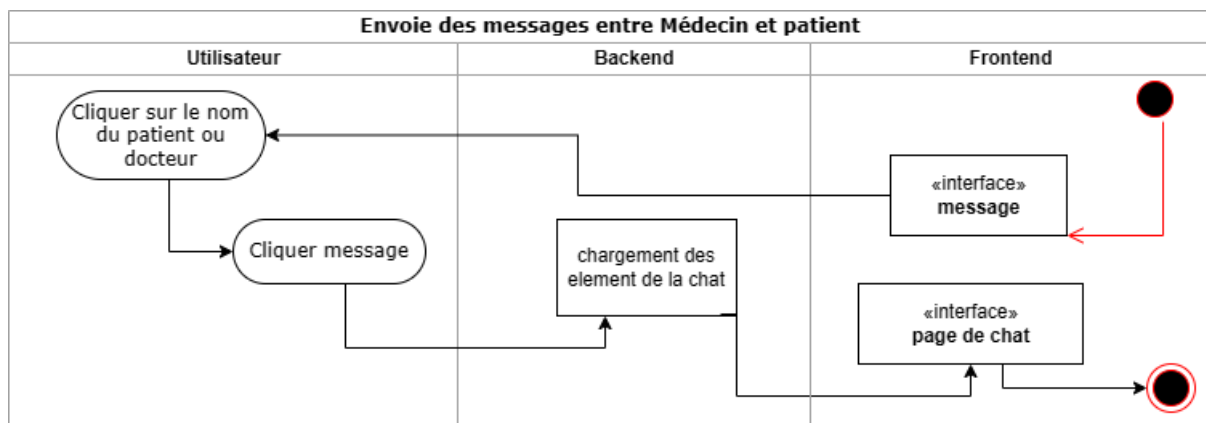


Figure 21: Diagramme d'activité pour s'envoyer des messages dans SantePlus

Diagramme d'activité pour le cas d'utilisation 11 –se déconnecter.

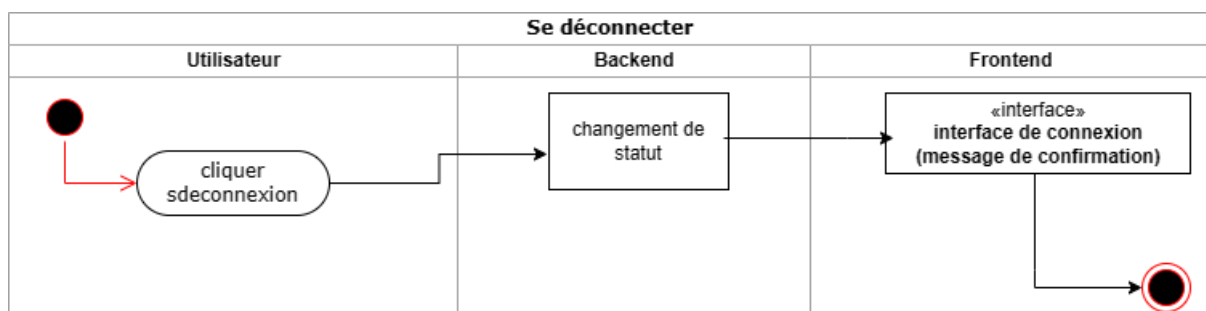


Figure 22: Diagramme d'activité pour se déconnecter de SantePlus

Comme on peut le constater, les diagrammes d'activité sont un outil précieux pour la modélisation du comportement des applications web. Ils permettent de visualiser la séquence des actions qui doivent être effectuées par l'utilisateur et par le système, ainsi que les données

qui sont échangées entre eux. Cette liste n'est pas exhaustive pour notre application web mais une présentation de la manière d'aborder ce genre des diagrammes.

### 3.4.5. Modélisation de la base de données.

Pour concevoir un système informatique, une des grandes parties consiste à modéliser les données. C'est en d'autres termes l'analyse et la conception de l'information dans le système. Cette partie consiste à identifier les entités logiques et dépendances entre ces entités.

Pour le cas de ce travail, nous utilisons une base de données SQLITE3 disponible dans un projet Django de python. Mais comme ce genre de base de données n'accepte pas les « array » et les « listes », une migration vers une base de données postgres a été nécessaire. La figure ci-dessous montre les différentes entités de la base de données utilisée.

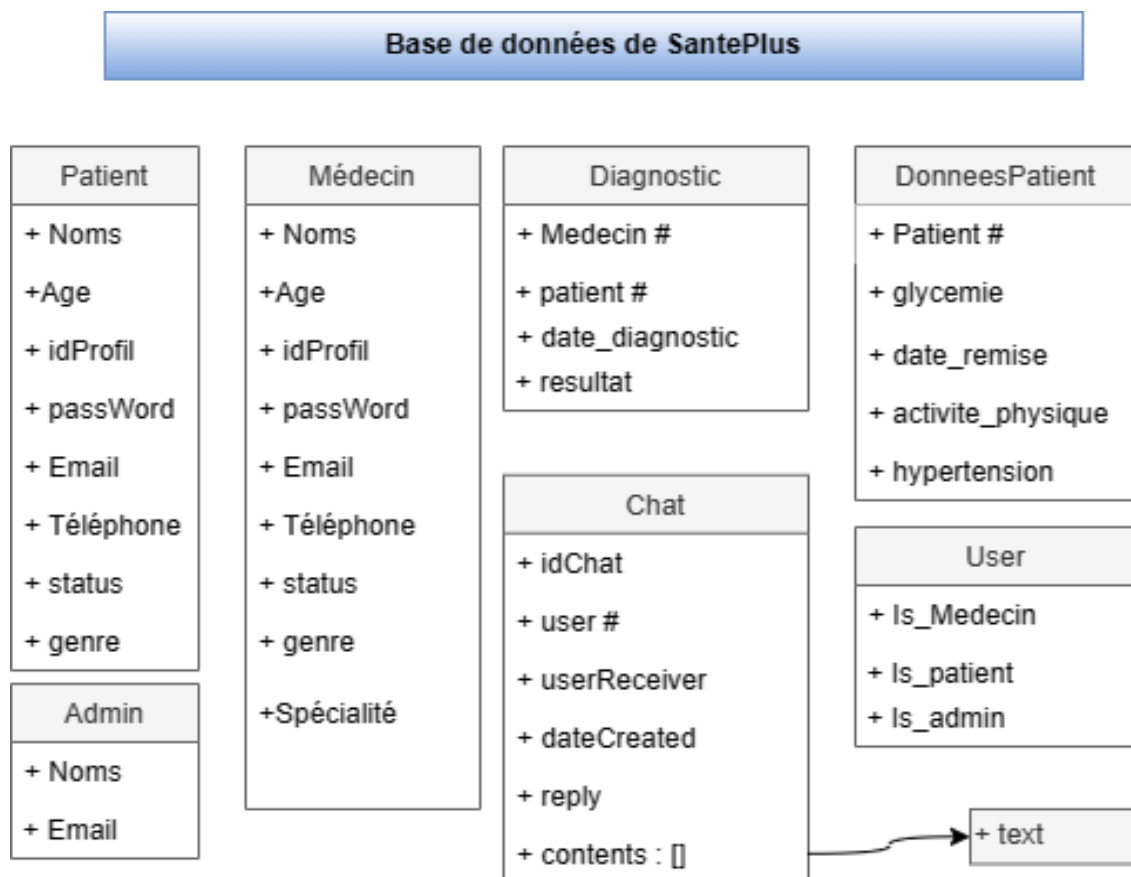


Figure 23: Modélisation de la base de données de l'application.

## Technologies utilisées pour réaliser l'application

Cette partie est réservée aux environnements utilisés pour la réalisation de ce projet.

### (a) Environnement matériel

L'équipement utilisé pour la réalisation de ce travail est composé d'un ordinateur portable avec un système d'exploitation Windows 10 Professionnel version 19045 de 64 bits. Les configurations sont les suivantes :

- Intel(R) Core(TM) i5-4300U CPU @ 1.90GHz 2.49 GHz, 2 Core(s), 4 Logical Processor(s)
- 8Go de mémoire vive.
- 500Go d'espace disque.

### Environnement logiciel

Pour développer notre logiciel, nous avons utilisés plusieurs logiciels et librairies.

Pour l'intelligence artificielle on peut citer :

- **Google Colaboratory** : couramment appelé "Google Colab", est un produit de Google Research. Colab permet à n'importe qui d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. C'est un environnement particulièrement adapté à la machine Learning, à l'analyse de données et à l'éducation. Ce service ne nécessite aucune configuration et permet d'accéder sans frais à des ressources informatiques, dont des GPU (Graphics Processings Units). [28]
- **Kaggle (pour le dataset)** : Kaggle est une plateforme web qui permet aux data scientists et aux professionnels de l'apprentissage automatique de se connecter, d'apprendre et de collaborer. Elle propose une variété de fonctionnalités et de services, mais dans ce travail cette plateforme a fourni une grande collection de jeux de données open source qui ont rapport avec le diabète et qui a servi de dataset.

Pour les API (Application Programming Interface) :

- **Le Framework Django** : Django est un Framework de Python qui encourage le développement rapide et une conception propre et pragmatique. Construit par des développeurs expérimentés, il s'occupe d'une grande partie du développement web, afin que vous puissiez vous concentrer sur l'écriture de votre application sans avoir besoin de réinventer la roue. Ce Framework est gratuit et open source. [29]

- **Visual studio code** : Visual Studio Code, souvent raccourci en VS Code, est un éditeur de code gratuit et open-source développé par Microsoft pour Windows, Linux et macOS. Il s'agit d'un éditeur léger mais puissant qui prend en charge une variété de langages de programmation et de frameworks, tels que Python, JavaScript, TypeScript, Java, C++, et C#. [30]

#### 3.4.6. Langages de programmation utilisé

Pour réaliser l'application web de diagnostic du diabète et de suivi du patient diabétique nous avons utilisé deux langages principaux : JavaScript et Python.

- **JavaScript (JS)** : JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives et à ce titre est une partie essentielle des applications web. Avec les langages HTML et CSS, JavaScript est au cœur des langages utilisés par les développeurs web. Une grande majorité des sites web l'utilisent, et la majorité des navigateurs web disposent d'un moteur JavaScript pour l'interpréter [31].
- **Python** est un langage de programmation de haut niveau à usage général. Sa philosophie de conception met l'accent sur la lisibilité du code avec l'utilisation d'une indentation significative [32].

Il est simple et a un grand nombre d'applications et de développements. Python est un langage de programmation populaire et recommandé dans le monde de l'intelligence artificielle, il est devenu le langage leader pour l'intelligence artificielle.

#### 3.4.7. Base de données.

Pour l'implémentation physique de la base de données, nous avons utilisé SQLITE.

Django fournit un système de gestion de base de données (SGBD) intégré qui facilite la création, la gestion et l'utilisation de bases de données. Ce SGBD est basé sur le modèle relationnel, qui est le modèle de base de données le plus courant.

Pour utiliser cette base de données dans Django, il faut d'abord configurer la connexion à la base de données dans le fichier **settings.py** de l'application. Ce fichier contient les paramètres de connexion à la base de données, tels que le nom de la base de données, le nom d'utilisateur, le mot de passe et le type de base de données.

La structure des tables de la base de données est définie par les classes Python du fichier **models.py**. Chaque modèle de données correspond à une table de la base de données.

Ici-bas une image qui visualise notre base de données dans le compte Superadmin de l'application Django.

AUTHENTICATION AND AUTHORIZATION	
Groups	<a href="#">+ Add</a> <a href="#">✎ Change</a>

DIAPREDAPP	
Diagnostics	<a href="#">+ Add</a> <a href="#">✎ Change</a>
Donnees patients	<a href="#">+ Add</a> <a href="#">✎ Change</a>
Medecins	<a href="#">+ Add</a> <a href="#">✎ Change</a>
Patients	<a href="#">+ Add</a> <a href="#">✎ Change</a>
Users	<a href="#">+ Add</a> <a href="#">✎ Change</a>

**Figure 24: Visualisation de la base de données dans le compte superadmin**

#### 3.4.8. Entraînement du modèle

Pour faire l'entraînement d'un modèle d'intelligence artificiel il est recommandé de passer par :

- **Préparation des données** : cette étape consiste à collecter les données d'entraînement, à les nettoyer et à les préparer pour l'entraînement. Cela peut inclure des tâches telles que la suppression des valeurs aberrantes, l'imputation des valeurs manquantes et la normalisation des données. Pour le cas de notre dataset :
  - Pas de valeurs qui sont nulles comme le montre la figure 25:

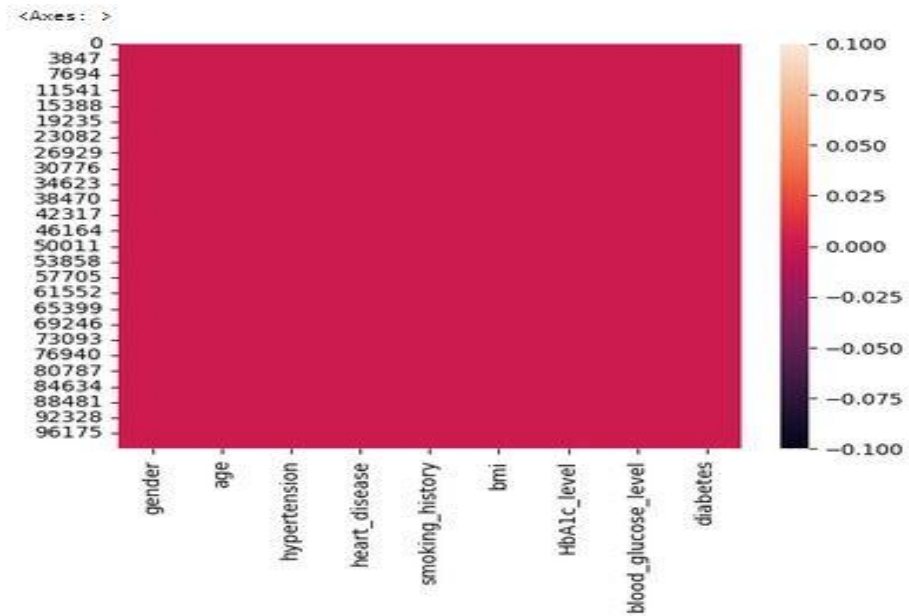


Figure 25: visualisation des valeurs nulles du dataset

- Quelques valeurs aberrantes sont présentes dans le dataset comme le montre la figure 27.

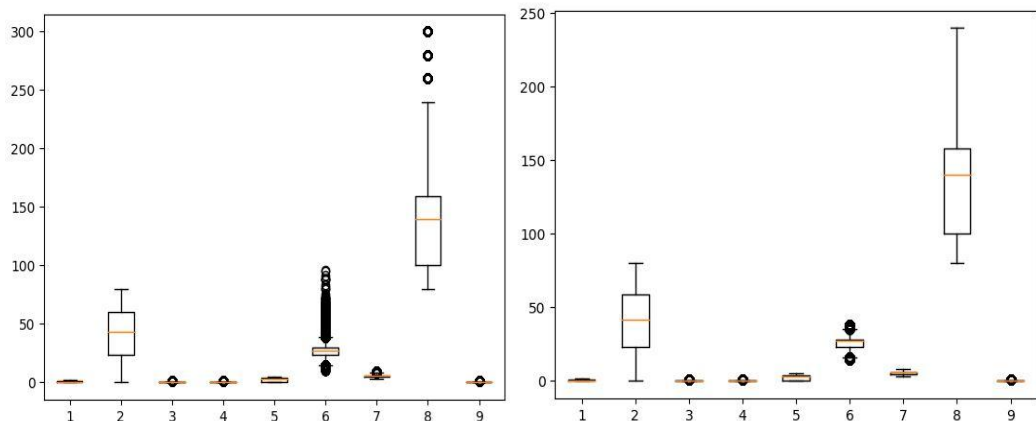


Figure 26: visualisation des valeurs aberrantes du dataset et leur élimination.

Après élimination des valeurs aberrantes 8113 lignes ou tuples sur 100000 sont supprimées faute de la présence des valeurs aberrantes pour les colonnes « bmi », « blood\_glucose\_level » et « HbA1c ».

La matrice de corrélation des données de ce dataset est comme suit :

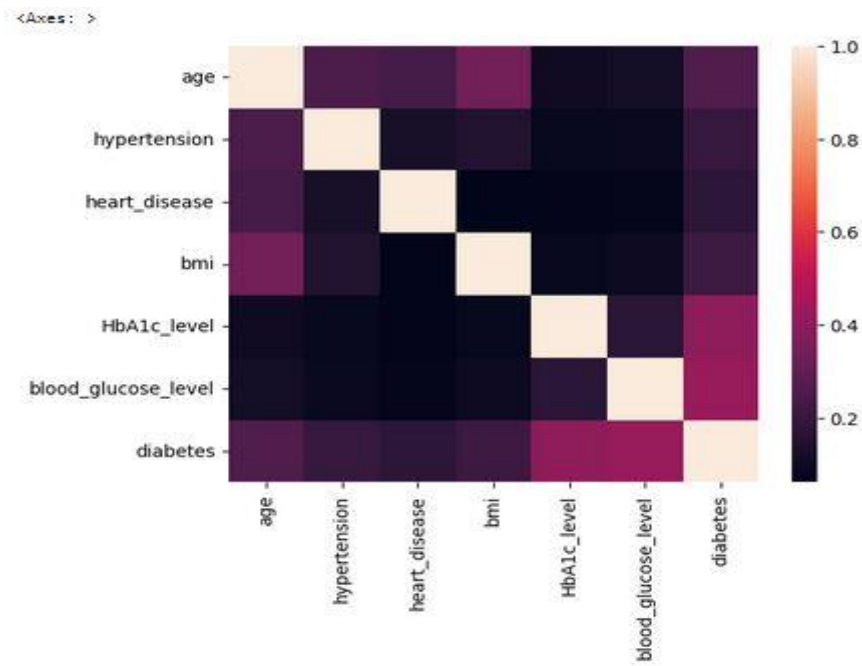


Figure 27:matrice de corrélation de différentes données du dataset

- **Sélection du modèle** : cette étape consiste à choisir le modèle d'apprentissage automatique qui convient le mieux à la tâche à effectuer. Pour ce travail c'est le modèle de régression.
- **Initialisation des paramètres** : cette étape consiste à initialiser les paramètres du modèle à des valeurs aléatoires ou à des valeurs spécifiques. Les paramètres du modèle sont les variables qui contrôlent son comportement.
- **Entraînement du modèle** : cette étape consiste à présenter le modèle aux données d'entraînement et à ajuster ses paramètres jusqu'à ce qu'il apprenne à effectuer la tâche donnée. L'entraînement du modèle a été effectué à l'aide d'un algorithme d'apprentissage automatique, la régression logistique.
- **Évaluation du modèle** : cette étape consiste à évaluer les performances du modèle sur des données de test. Cela permet de s'assurer que le modèle fonctionne bien sur des données qu'il n'a pas vues auparavant.

### 3.5. Les interfaces du système

Il a été mentionné dans les lignes qui précèdent que notre application intègre une interface homme-machine permettant une interaction entre les utilisateurs de l'application SantePlus et l'application en soit. Les captures d'écran suivantes donnent un aperçu général des pages de notre application web.

### 3.5.1. Page d'authentification

Cette interface constitue le point d'entrée de notre application, elle permet aux utilisateurs du système de se connecter en utilisant leurs noms d'utilisateur et leurs mots de passe.



**Figure 28: login page de l'application SantePlus**

### 3.5.2. Page d'enregistrement

Si l'utilisateur ne possède pas encore de compte, il est invité à contacter l'hôpital en suivant les instructions fournies sur la page de connexion. La page suivante lui permettra alors de créer un compte utilisateur dans la base de données, lui permettant ainsi d'accéder aux services de l'application SantePlus. Le formulaire doit être rempli soit par le médecin, soit par l'administrateur, avec une distinction de rôles : l'administrateur peut spécifier un rôle, tandis que le médecin peut créer un compte avec uniquement le rôle de patient.

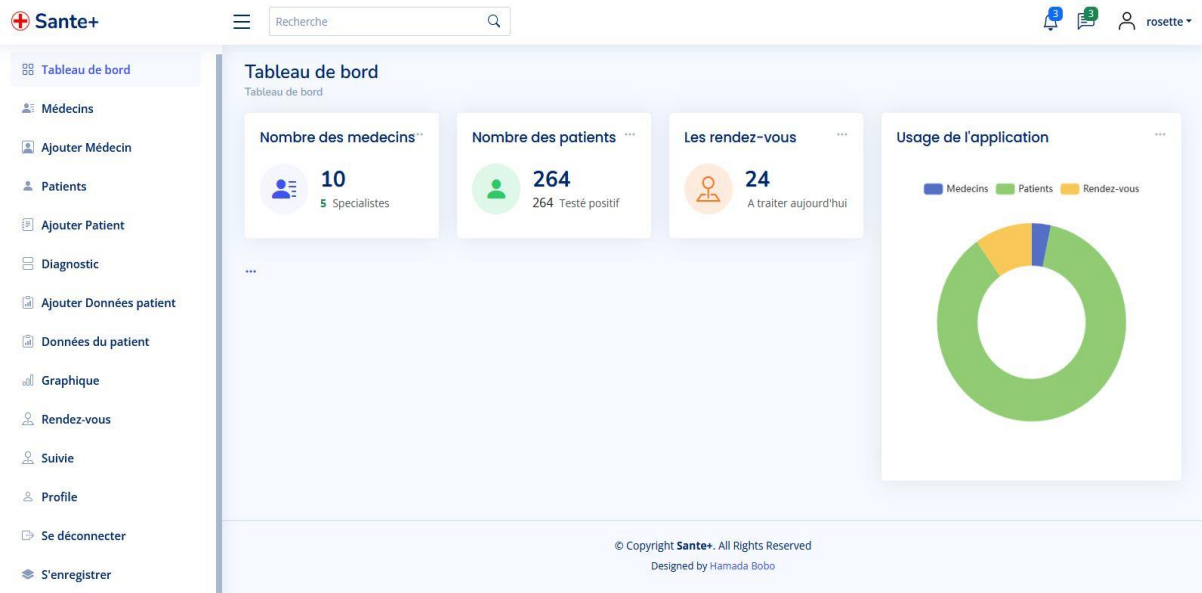
The image shows a web form for creating a user account. At the top, there is a logo with a red cross and the text 'Sante+'. Below the logo, the title 'Créer un compte' is centered. Underneath the title, a subtitle reads 'Entrer les informations personnelles pour créer un compte utilisateur'. The form is divided into sections: 'Login page' with a 'Noms' field; 'Email' with two stacked input fields; 'Mot de passe' with one input field; and 'confirmer le Mot de passe' with another input field. At the bottom left, there are three radio button options: 'Admin', 'medecin', and 'patient'. A blue button labeled 'Créer un compte' is positioned at the bottom center. Below the button, there is a link that says '!!elle a déjà un compte? Diagnostiquer'.

**Figure 29: Page d'enregistrement des utilisateurs**

Comme dit précédemment, un utilisateur peut se connecter en tant que médecin, en tant qu'admin ou comme patient. Si on y va cas par cas, voici les interfaces selon l'utilisateur.

### 3.5.3. Tableau de bord pour un utilisateur administrateur

Une fois connecté comme administrateur, l'utilisateur a tous les droits. Il peut accéder à toutes les interfaces de l'application et utiliser toutes les fonctionnalités qui sont : voir la liste des médecins et celle des patients, ajouter un médecin et/ou un patient, faire un diagnostic, ajouter les données d'un patient et les visualiser, voir les différents profils des utilisateurs et faire leur suivi, programmer un rendez-vous. La figure 31 ci-dessous est l'illustration du tableau de bord.

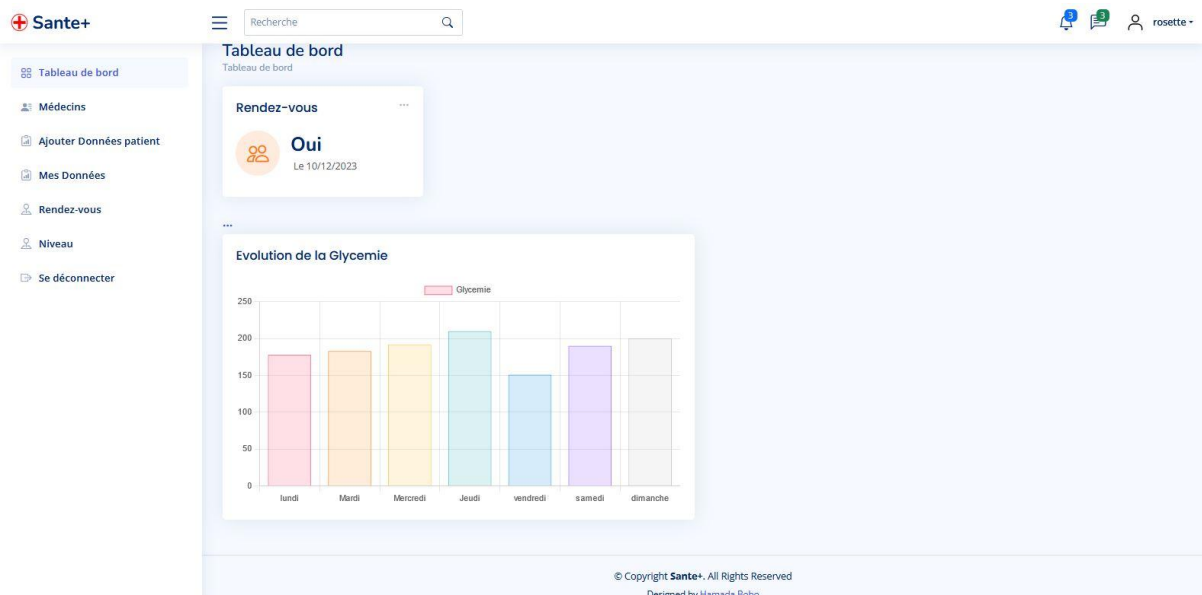


**Figure 30: admin dashboard ou tableau de bord de l'administrateur**

Le médecin a presque le même tableau de bord sauf que pour son cas la fonctionnalité pour ajouter les médecins ne lui est pas donnée.

### 3.5.4. Tableau de bord pour un utilisateur Patient.

Surtout que les utilisateurs patients ne sont pas très informés sur l'application, cet outil leur offre un tableau de bord simple qui ne leur donne accès qu'aux fonctionnalités qui les concernent. Cette capture en parle mieux :



**Figure 31: Tableau de bord pour un patient**

### 3.5.5. Page médecins.

Cette page permet de voir la liste des médecins, leurs spécialités respectives et leurs contacts téléphoniques et donne la possibilité de les contacter.

**Liste des médecins et leurs spécialités**  
Home / Medecin

Les médecins

10 entries per page

#	Noms	Email	Specialite	Telephone	Action
1	Bienvenue	mujos@gmail.com	Neurochirurgier	+243975851556	<a href="#">Contacter</a>
2	Amani hamada	amani45amadabonheur@gmail.com	pediatre	+256788808723	<a href="#">Contacter</a>
3	Bibishe Ham	bonheur@gmail.com	doctorlove	+243991904338	<a href="#">Contacter</a>

Showing 1 to 3 of 3 entries

Figure 32:page de visualisation des médecins

### 3.5.6. Ajouter Médecin/Patient

Cette page est un formulaire qui permet d'ajouter un médecin/patient dans le système. Ce processus n'est possible que si le sujet est déjà un utilisateur.

**Formulaire pour Ajouter un medecin**  
Home / Medecin / Ajouter un medecin

Elements importants

User\*

Noms\*

Specialite\*

Email\*

Telephone\*

[Ajouter](#)

Figure 33: Formulaire pour ajouter un médecin

Un autre formulaire presque similaire à celui-ci est présent dans l'application pour ajouter un patient sauf que dans ce cas les inputs sont les informations propres aux patients. Une autre

page pour visualiser, contacter et programmer un rendez-vous est disponible dans l'application.

### 3.5.7. Diagnostic

La page de diagnostic contient une partie pour le formulaire à remplir avec les informations d'après les résultats du laboratoire et une autre partie pour visualiser le résultat de diagnostic et le degré de précision. La figure ci-dessous le montre clairement.

**Diagnostic**  
Home / Diagnostics / Diagnostics

Entrer ces informations Après test

Genre: Masculin | Age: 50.0

Hypertension: Oui | Maladie cardiaque: Oui | Historique tabagique: Oui

BMI: 27.32 | Taux d'HbA1c: 6.6 | Glycemie: 140.0

**Diagnostic**

Résultat de diagnostic:  
**Positif**

Précision  
**74.18652886819105 %**

temps d'exécution:  
**1.2497131824493408 S**

**Figure 34:page de diagnostic**

### 3.5.8. Suivi du patient

Une fois que pour le patient, le résultat de diagnostic est « positif », le suivie consiste en la réception des données provenant d'un patient X, recevoir une notification dans le compte d'un médecin lorsqu'il y a anomalie dans les valeurs envoyées par le patient mais aussi le contacter et programmer un rendez-vous. Ici-bas le formulaire que le patient doit utiliser pour envoyer ses données.

**Formulaire pour Ajouter les donnees du patient**  
 Home / Medecin / Ajouter les donnees du Jour

**Les donnees d'aujourd'hui**

Patient\*

Glycemie\*

Hypertension\*

Medicament\*

Activites physique\*

Figure 35 : formulaire pour ajouter les données du jour.

### 3.5.9. Données du patient

Cette page permet de visualiser un diagramme qui trace l'évolution des données du patient juste par un clic sur le bouton « suivre » de la liste des patients ou en passant par l'aside bar ou le menu « suivi ». L'image suivante montre le diagramme des données d'un des patients.

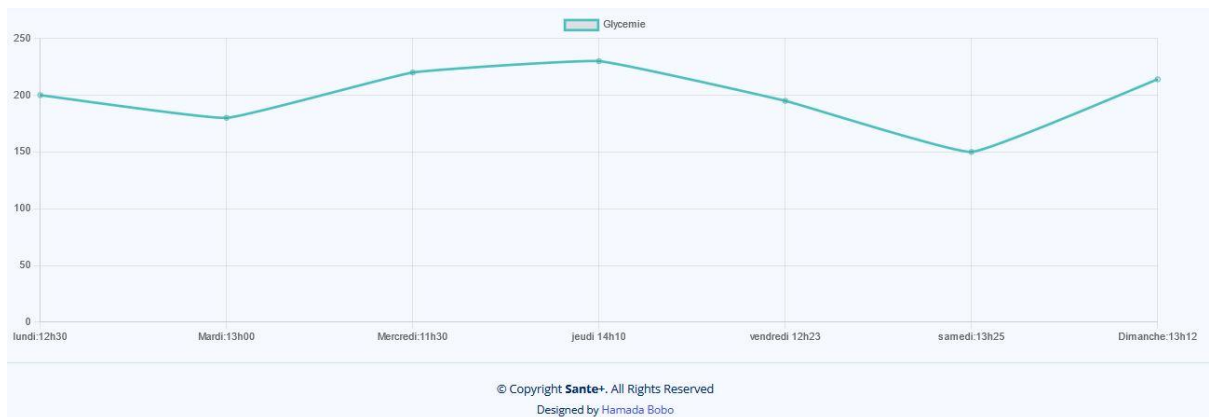
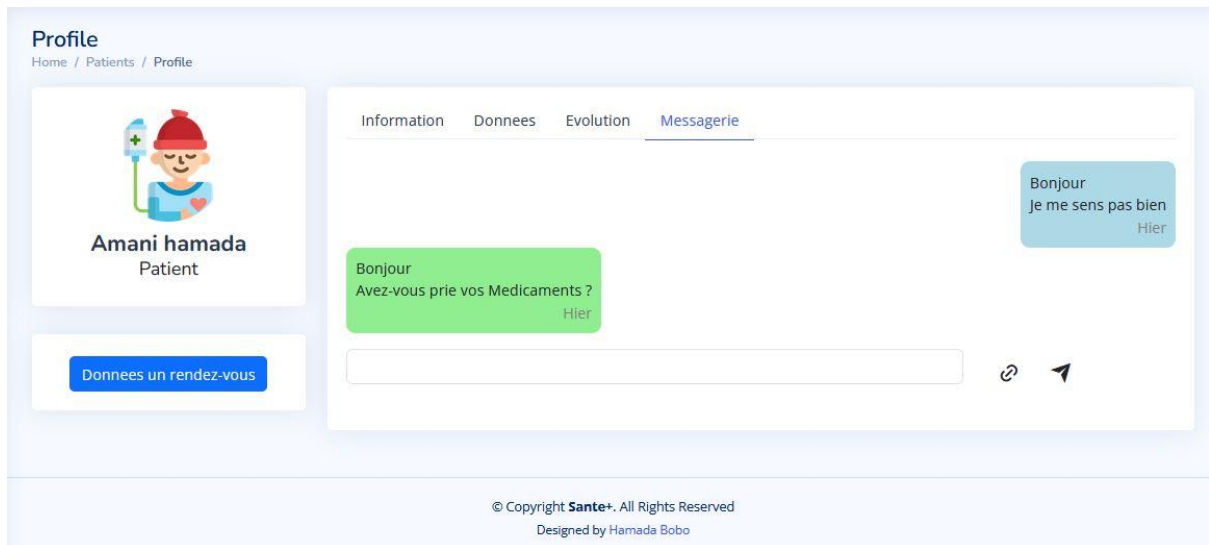


Figure 36:visualisation des données d'un patient

### 3.5.10. Contacter

Un patient peut contacter un médecin pour telle ou telle autre raison et un médecin peut contacter le patient si les données présentent une certaine anomalie. Cette application donne la possibilité de le faire. La figure 37 le montre :



**Figure 37: contacter un utilisateur**

### 3.6. Conclusion partielle

Dans ce chapitre, nous sommes partie de la conception de l'application vers la présentation des technologies utilisées (langages, outils utilisés) pour la réalisation l'interface et la base de données physique de notre application. Cette partie nous a aidé aussi à passer des modèles d'intelligence artificielle pour diagnostic du diabète en suivant toutes ses étapes notamment : acquisition des données, prétraitement, définition de modèle, l'entrainement et validation puis l'évaluation de modèle, vers les interfaces du système en donnant une description sur leur utilité.

## Conclusion générale

---

Dans ce travail, rédigé dans le cadre de la fin de nos études, nous avons visé mettre en œuvre un outil d'aide au diagnostic du diabète et qui fait le suivi du patient diabétique. Basé sur la régression logistique, une des méthodes de l'intelligence artificielle, cet outil permet de réaliser plus rapidement un diagnostic, à enregistrer certaines données liées au patient dans une base de données pour le suivi d'un patient par un médecin et faciliter un médecin non spécialiste en diabète.

Ce travail répond aux questions de recherches que nous nous sommes posées dans l'introduction :

- Serait-ce possible de réduire le temps de diagnostic du diabète tout en ayant une haute précision ?
- Comment utiliser les nouvelles technologies comme un moyen pour faciliter le suivi du patient ?

Cette application web permet d'enregistrer des données liées au patient et ce qui facilite le suivi d'un patient par un médecin. Dans ce travail nous avons présenté la télémédecine, le diabète et la manière dont son diagnostic peut se faire en utilisant l'apprentissage automatique en particulier ou l'intelligence artificielle en générale. Les concepts de l'apprentissage automatique sont présents dans ce travail et sont mis application pour arriver à faire le choix du modèle à utiliser.

Avec une précision de 97,67% la régression logistique a été choisie sur trois méthodes d'entraînement utilisées. Pour parvenir aux résultats obtenus, nous avons subdivisée notre travail en trois chapitres et chacun de ce chapitre avait un output utile pour le chapitre suivant. Le premier chapitre parle des concepts de base sur le diabète et sur la télémédecine pour comprendre la maladie et avoir une idée sur comment la technologie intervient en médecine. Le chapitre deux parle de l'apprentissage automatique pour expliquer comment fonctionne l'intelligence artificielle dans la prédiction de certains cas. En fin le chapitre trois qui nous a servi de la présentation des résultats. Parlant des avantages de l'outil implémenté, nous pouvons dire que cet outil est utile :

- Aux médecins dans leur dépistage et suivi médical ;
- Aux patients pour recevoir des orientations de la part des médecins.
- Aux étudiants en médecine et aux personnels soignants qui ne sont pas spécialistes.

- Aux étudiants en technologies qui veulent continuer avec ce projet, en continuant les entraînements de nouvelles données ; et en y ajoutant des nouvelles fonctionnalités.

Nous ne prétendons pas avoir conçu un modèle parfait et moins encore avoir touché toute la matière concernant ce sujet. Ainsi toute personne désirant continuer ces recherches dans ce domaine peut suivre les orientations suivantes : Intégrer les notifications dans l'application et faire que la conversation soit interne pour ne plus utiliser d'autres applications externes pour communiquer.



## Références

---

- [1] Université de Stanford, «Prédiction du diabète de type 2 à partir de données cliniques et génomiques,» *Nature Medicine*, p. 559, 2018.
- [2] E. d. d. d. G. d. web, «Application mobile,» chez *Glossaire du web*, 2023, p. 1384.
- [3] RanjitPal Singh, *Qu'est-ce qu'une application native ? Un guide complet*, p. 2, 2023.
- [4] B. e. CHEKLAT.L, «Conception et réalisation d'une application mobile sensible au contexte pour un musée,» Juillet 2013..
- [5] L. G. e. P. Wilson, «Mobile Application Development,» p. 232, 2011.
- [6] Association médicale canadienne (CMAJ) , «Télémédecine : une révolution dans les soins de santé,» 2017.
- [7] Parlement français, Code de la santé publique.
- [8] Haute Autorité de santé (HAS), Guide de bonnes pratiques pour la téléconsultation et la téléexpertise, 2022.
- [9] J.-L. D. e. J.-F. Contandriopoulos, *Télémédecine : un nouveau paradigme de soins*, 2017.
- [10] A.-M. M. e. M.-C. d. Peretti, *Le diabète, une épidémie silencieuse*, éditions Armand Colin, 2013.
- [11] J.-M. B. e. Jean-Pierre, *Diabète, comprendre et agir*, Malakoff, en France: éditions Dunod, 2015.
- [12] H. D, Thèse de Doctorat: Apprentissage Automatique dans un Agen, Constantine: université Mentouri, 2014.
- [13] A. M.Y, Thèse de doctorat: Mise en oeuvre de réseaux de neurones pour la modélisation de cinétique, Tunis, 2007.
- [14] J. Brownlee, *Master Machine Learning Algorithms*, Academia, 2016.

- [15] L. M, «. Mémoire de fin d'etude:Anticipation des changements des notes des obligations des portefeuille d'un assureur,» Universtés Dauphine, paris, 2017.
- [16] F. B. e. D. E. Pierre Geurts, Apprentissage automatique, Heidelberg,: Springer Verlag, 2019.
- [17] T. M. Mitchell, Introduction à l'apprentissage automatique, 2015.
- [18] S. R. e. P. Norvig, ntelligence Artificiel: a modern approach, New york: Prentice Hall, 1995.
- [19] Shrivastav, «Everything You Need To Know About Decision Trees (With Code),» [En ligne].
- [20] java T, «Random Forest Algorithm.,» [En ligne]. Available: <https://www.javatpoint.com/machine-learning-random-forest-algorithm>.
- [21] Tutorials point., «Classification Algorithms - Random Forest,» [En ligne].
- [22] CNIL, «Apprentissage automatique,» [En ligne]. Available: [https://www.cnil.fr/fr/definition/apprentissage-automatique#:~:text=%23Intelligence%20artificielle%20%28IA%29%20L%E2%80%99apprentissage%20automatique%20%28machine%20learning%20en,%C3%A0%20partir%20de%20donn%C3%A9es%2C%20via%20des%20mod%C3%A8les%20math%C3%](https://www.cnil.fr/fr/definition/apprentissage-automatique#:~:text=%23Intelligence%20artificielle%20%28IA%29%20L%E2%80%99apprentissage%20automatique%20%28machine%20learning%20en,%C3%A0%20partir%20de%20donn%C3%A9es%2C%20via%20des%20mod%C3%A8les%20math%C3%99).
- [23] Canadian family physician, «The official journal of the college of family physicians of canada,» [En ligne]. Available: <https://www.cfp.ca/content/65/6/e251/tab-figures-data>.
- [24] K. P.Murphy, Machine Learning: A Probabilistic Perspective, cambridge: MIT press, 2012.
- [25] DellaData, «Introduction à la régression logistique,» 2022. [En ligne]. Available: <https://delladata.fr/regression-logistique/>. [Accès le 2023].
- [26] N. W. e. A. S. M. MA.Kamal, Prediction of Diabetes Using Machine Learning Algorithms in Healthcare., 2018.
- [27] Université de Lorraine, «Diagrammes d'activité,» chez *UML Diagrammes*, Nancy,

Université de Lorraine, 2018, p. 2.

- [28] Google, «Google Colab,» [En ligne]. Available: <https://research.google.com/colaboratory>. [Accès le 20 11 2023].
- [29] GOOGLE, «Commencez avec Django,» [En ligne]. Available: <https://www.djangoproject.com/>. [Accès le 20 11 2023].
- [30] wikipedia, «Visual Studio Code,» 10 2023. [En ligne]. Available: [https://fr.wikipedia.org/wiki/Visual\\_Studio\\_Code](https://fr.wikipedia.org/wiki/Visual_Studio_Code). [Accès le 20 11 2023].
- [31] D. A. Rauschmayer, JavaScript for impatient programmers, ECMAScript, 2022.
- [32] A. Shaw, LEARN PYTHON THE HARD WAY Third Edition, New York: Safari Inc., 2014.

## Annexe

---

### Définition du modèle

```
def result(request):

    # df = pd.read_csv(r'C:\Users\hamada\Desktop\memoire\diabetes_prediction_dataset.csv')

    df = pd.read_csv(r'C:\Users\hamada\PycharmProjects\diabetePrediction\static\diabetePrediction\dataset\diabetes_prediction_dataset.csv')

    # from sklearn.preprocessing import LabelEncoder

    encoder = LabelEncoder()

    columns_to_encode = ['gender', 'smoking_history']

    for column in columns_to_encode:

        df[column] = encoder.fit_transform(df[column])

    df_complete = df.copy()

    X = df_complete.drop('diabetes', axis=1) # dropping the outcome variable from features
(X)

    y = df_complete.diabetes # redefining the outcome variable as y

    # Separating the dataset into training and validation sets

    X_train, X_valid, y_train, y_valid = train_test_split(X, y, test_size=0.2, random_state=42)

    # Simple XGBoost model

    diabetes_model_1 = LogisticRegression()

    diabetes_model_1_train = diabetes_model_1.fit(X_train, y_train)
```

```
val1 = float(request.GET["n1"])
val2 = float(request.GET["n2"])
val3 = float(request.GET["n3"])
val4 = float(request.GET["n4"])
val5 = float(request.GET["n5"])
val6 = float (request.GET["n6"])
val7 = float (request.GET ["n7"])
val8 = float(request.GET["n8"])

pred = diabetes_model_1.predict([[val1, val2, val3, val4, val5, val6, val7, val8]])

result1 = ""
if pred == [1]:
    result1 = "positif"
else:
    result1 = "negatif"
return render(request, "DiaPredApp/page-predict.html", {"result2": result1})
```

## Structure de dossiers

