RÉPUBLIQUE DU BÉNIN
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ D'ABOMEY-CALAVI

INSTITUT DE FORMATION ET DE RECHERCHE EN INFORMATIQUE

BP 526 Cotonou Tel : +229 55 02 80 70
https://ifri-uac.bj/ Courriel : contact@ifri.uac.bj

MÉMOIRE

pour l'obtention du

Diplôme de Licence en Informatique

Option : Génie Logiciel

Présenté par :
Fréjus **LAWSON**

Conception d'une Application de Prédiction du Diabète

Sous la supervision :
Dr Pierre Jerome **ZOHOU**

Année Académique : **2022-2023**

Sommaire

Dédicace ii

Remerciements iii

Résumé iv

Abstract v

Table des figures vi

Liste des tableaux vii

Liste des acronymes viii

Introduction 1

1 État de l'art 4

2 Méthodologie de travail et Conception du système 13

3 Résultats et discussions 35

Conclusion 43

Bibliographie 45

Webographie 46

Table des matières 47

Dédicace

Je dédie ce travail à :

- L'Eternel, mon Dieu, pour son soutien et sa protection tout au long de mon cursus académique qui s'achève ;
- Mon père Dorothé LAWSON ;
- Ma mère Reine KPANOU ;
- Mes sœurs Dorell et Magnificat ;
- Mon frère Fiat et tous ceux ayant contribué d'une manière ou d'une autre à la réalisation de ce travail.

Remerciements

J'adresse mes sincères remerciements

- Au Professeur C. Eugène EZIN, Directeur de l'Institut de Formation et de Recherche en Informatique (IFRI) pour tout ce qu'il met en oeuvre pour la réussite de l'Institut et de ses étudiants ;
- Au Professeur Gaston G. EDAH, Directeur Adjoint de l'Institut de Formation et de Recherche en Informatique (IFRI) pour tout le travail et l'accompagnement le long de la formation des étudiants ;
- A tout le corps professoral de l'Institut de Formation et de Recherche en Informatique pour l'accompagnement et la formation de pointe durant ces trois dernières années ;
- Au Dr Pierre Jérôme ZOHOU, mon maître de mémoire, qui, malgré ses nombreuses occupations, a accepté de superviser ce travail ;
- A Monsieur Ange KLIKA, mon tuteur de stage, pour ses conseils judicieux et son encadrement à l'insertion professionnel ;
- A mes amis Basile, Ludel, Miguel et Stephen je vous dis merci .

Résumé

Ce mémoire de fin de formation propose une solution aux défis rencontrés par les professionnels de la santé dans le diagnostic précoce et le suivi du diabète. Nous avons développé une application web nommée DiabOracle qui permet à toute personne de saisir ses données cliniques et de recevoir des prédictions sur son risque de diabète. Cette application aide également les professionnels de la santé à surveiller la prévalence du diabète au sein de la population et à suivre leurs patients. Pour réaliser cette solution, nous avons utilisé la régression logistique sur un ensemble de données contenant 384 enregistrements, obtenant une précision de 93.48% pour le modèle de prédiction. Nous avons également employé le langage de modélisation UML pour créer des diagrammes de cas d'utilisation, de classes et de séquences, ainsi que le langage de programmation Python et le framework Django, avec MySQL comme moteur de base de données.

Mots clés : diabète, prédiction, application web, régression logistique, UML, Python, Django, MySQL

Abstract

This final project thesis proposes a solution to the challenges faced by healthcare professionals in diagnosing and monitoring diabetes. We have developed a web application named DiabOracle that allows anyone to enter their clinical data and receive predictions about their diabetes risk. This application also helps healthcare professionals to monitor the prevalence of diabetes within the population and to track their patients. To achieve this solution, we used logistic regression on a dataset containing 384 records, achieving an accuracy of 93.48% for the prediction model. We also used the UML modeling language to create use case diagrams, class diagrams, and sequence diagrams, as well as the Python programming language and the Django framework, with MySQL as the database engine.

Key words: Diabetes, prediction, web application, logistic regression, UML, Python, Django, MySQL

Table des figures

1.1 DiabTrend . . . 5
1.2 Dashboard de l'application One Drop . . . 6
1.3 Page d'acceuil de l'application Diabet AI . . . 7
1.4 Page de prédiction de l'application MD Calc . . . 8
1.5 Page de prédiction de l'application Diabetes Australia Risk Calculator . . . 9

2.1 Diagramme de cas d'utilisation . . . 17
2.2 Diagramme de classes . . . 19
2.3 Diagramme de séquences : faire une prédiction . . . 20
2.4 Extrait des données brutes collectées . . . 23
2.5 Chargement des données et resultats . . . 25
2.6 Description des données avant nettoyage . . . 26
2.7 Description des données après nettoyage . . . 27
2.8 Entrainement du modèle . . . 28
2.9 Score de précision du modèle . . . 29
2.10 Constantes du risque P . . . 30
2.11 Fonction pour impléménter le risque P . . . 31

3.1 Page D'accueil . . . 36
3.2 Informations personnelles . . . 37
3.3 Présentation des résultats sous forme de Diagramme . . . 37
3.4 Dashbord . . . 38
3.5 Enrégistrement des constantes . . . 39
3.6 Historique des prédictions . . . 40

Liste des tableaux

1.1 Avantages et inconvénients des solutions de gestion du diabète existantes 10

Liste des acronymes

APDP :

Autorité de Protection des Données Personnelles : C'est l'organisme chargé de veiller à l'application de ces lois et de protéger les droits des individus en matière de données personnelles au Bénin3. Elle s'assure que les données personnelles sont traitées de manière légale, équitable et transparente, et que les droits des personnes concernées sont respectés viii, *Glossaire :* APDP

API :

Application Programming Interface est une interface de programmation qui permet à des applications de communiquer entre elles et d'échanger des données ou des services. Les API fournissent un ensemble de fonctions et de protocoles qui facilitent l'interaction entre différents logiciels, en cachant les détails de leur implémentation viii, *Glossaire :* API

CSS :

Cascading Style Sheets, est un langage de feuille de style utilisé pour décrire la présentation des documents HTML. Il permet de contrôler l'apparence des éléments sur une page web, comme les couleurs, les polices, les espacements, et la disposition générale. viii, *Glossaire :* CSS

DCU :

Diagramme de Cas d'Utilisation: En génie logiciel, un DCU est un diagramme UML utilisé pour représenter le comportement fonctionnel d'un système logiciel. Il montre les interactions entre les utilisateurs (appelés acteurs) et le système à travers différents cas d'utilisation. viii, *Glossaire :* DCU

DIABETE :

Le diabète est une maladie chronique caractérisée par un taux élevé de glucose (sucre) dans le sang, aussi appelé hyperglycémie. Cela se produit lorsque le corps ne produit pas suffisamment d'insuline ou lorsque l'insuline produite ne fonctionne pas correctement. L'insuline est une hormone produite par le pancréas, qui permet au glucose de pénétrer dans les cellules pour être utilisé comme source d'énergie. viii, *Glossaire :* DIABETE

HTML :

"HyperText Markup Language, est un langage de balisage utilisé pour créer des pages web. Il permet de structurer le contenu d'une page en utilisant des éléments sémantiques et structurels,

tels que les titres, les paragraphes, les listes, les images, et bien plus12. HTML est essentiel pour définir comment le contenu doit être affiché dans un navigateur web. ix, *Glossaire :* HTML

IMC :

Indice de Masse Corporelle ix, *Glossaire :* IMC

iOS :

"iPhone Operating System", est le système d'exploitation mobile développé par Apple Inc. pour ses appareils mobiles, y compris les iPhone, les iPad et les iPod Touch1 ix, *Glossaire :* iOS

LAMPP :

LAMPP est un acronyme qui désigne un ensemble de logiciels utilisés pour créer et gérer des applications web. Il se compose des éléments suivants : Linux : Le système d'exploitation. Apache : Le serveur web. MySQL : Le système de gestion de base de données relationnelle. PHP : Le langage de script côté serveur. Perl : Un langage de programmation, souvent utilisé pour les scripts et les tâches d'administration système. ix, *Glossaire :* LAMPP

ORM :

Object-Relational Mapping est une technique de programmation qui crée un pont entre les bases de données relationnelles et la programmation orientée objet. En utilisant un ORM, les développeurs peuvent interagir avec une base de données en manipulant des objets plutôt qu'en écrivant directement des requêtes SQL ix, *Glossaire :* ORM

SGBDR :

Système de Gestion de Base de Données Relationnelle est un logiciel utilisé pour stocker, gérer, interroger et récupérer des données organisées en tables. Les bases de données relationnelles sont basées sur le modèle relationnel, où les données sont structurées en lignes et colonnes, et les relations entre les tables sont établies par des clés primaires et étrangères. ix, *Glossaire :* SGBDR

UML :

Le Langage de Modélisation Unifié, de l'anglais Unified Modeling Language (UML), est un langage de modélisation graphique à base de pictogrammes conçu comme une méthode normalisée de visualisation dans les domaines du développement logiciel et en conception orientée objet. ix, *Glossaire :* UML

Introduction Générale

Le diabète est l'une des maladies chroniques les plus courantes au niveau mondial, affectant des millions de personnes de tout âge, sexe et milieux socio-économiques. En raison des habitudes de vie sédentaires modernes et de la prévalence de l'obésité, les taux de diabète continuent d'augmenter, faisant de cette maladie un problème de santé publique majeur.

1. Problématique

La gestion et la prévention du diabète de type 2 représentent un défi majeur de santé publique, notamment en Afrique, où la maladie est souvent sous-diagnostiquée et mal suivie. Les méthodes actuelles de dépistage et de gestion du diabète, ainsi que les applications mobiles disponibles, présentent plusieurs limitations, ce qui diminue leur efficacité. Ces insuffisances soulèvent plusieurs problèmes :

- **Absence de conseils personnalisés** : Les applications de prédiction actuelles ne fournissent pas de conseils personnalisés ni de suivi médical adapté aux besoins individuels des utilisateurs, ce qui limite leur impact sur la prévention et la gestion du diabète.
- **Portée limitée aux utilisateurs individuels** : Ces solutions ne sont généralement pas conçues pour un usage clinique par les professionnels de santé, réduisant ainsi leur utilité pour un diagnostic médical complet et un suivi à long terme des patients diabétiques.
- **Accessibilité restreinte** : La plupart des applications ne sont pas multiplateformes, étant disponibles uniquement sur Android ou iOS. Cela limite l'accès des utilisateurs utilisant d'autres systèmes et freine l'intégration avec les systèmes d'information de santé existants.
- **Données non adaptées au contexte africain** : Les outils actuels se basent principalement sur des études et des données provenant d'autres régions du monde, ce qui réduit leur pertinence et leur précision pour la population africaine. Or, les facteurs génétiques et environnementaux spécifiques à cette région influencent considérablement l'évaluation du risque de diabète.
- **Suivi incomplet des facteurs de risque** : Les applications existantes ne permettent pas aux utilisateurs de suivre de manière dynamique l'évolution de leurs principaux facteurs de risque (poids, taille, IMC, etc.), ce qui est important pour une gestion proactive du diabète.

Ces différents points mettent en évidence la nécessité d'une solution plus adaptée et robuste, intégrant un suivi personnalisé, une accessibilité multiplateforme, et une validation basée sur les réalités africaines, afin d'améliorer la prévention et la gestion proactive du diabète dans la sous-région.

2. Contexte du Projet et justification

La détection précoce du diabète est important pour commencer le traitement à temps et prévenir les complications graves qui peuvent en découler. Toutefois, le diagnostic du diabète peut souvent être retardé en raison de l'absence de symptômes évidents aux stades initiaux de la maladie. C'est dans ce contexte qu'une approche de prédiction du diabète basée sur l'apprentissage automatique peut avoir un impact significatif.

Le développement d'une application de prédiction précise pour le diabète peut aider à identifier les individus à haut risque avant l'apparition des symptômes. En facilitant la détection précoce, ces individus peuvent bénéficier d'un suivi médical plus attentif et d'interventions de style de vie pour prévenir ou retarder le développement de la maladie. De plus, cela peut aider à économiser des ressources médicales en réduisant le nombre de personnes atteintes de complications graves du diabète nécessitant des soins médicaux intensifs.

3. Objectif du Projet et Description de l'Application

3.1 Objectif général

L'objectif principal de notre projet est de concevoir une application de prédiction du diabète de type 2 reposant sur un algorithme de machine learning capable d'évaluer le risque d'un patient de développer cette maladie. Cet algorithme analysera une série de paramètres de santé saisis par les utilisateurs, tels que l'âge, la glycémie, l'indice de masse corporelle (IMC) et le nombre de grossesses (pour les femmes), afin de fournir une prédiction précise et personnalisée du risque. Cette approche vise non seulement à permettre aux utilisateurs de mieux comprendre et gérer leur état de santé, mais aussi à renforcer la prévention et le suivi médical via une plateforme optimisée pour le contexte africain. L'application se positionne également comme un outil précieux pour les professionnels de la santé, contribuant à une meilleure gestion de la prévalence croissante du diabète et offrant aux individus un moyen efficace de surveiller leur propre santé.

3.2 Objectifs spécifiques du projet :

- **Collecte des données** : Développer un système de collecte de données utilisateur incluant des paramètres tels que l'âge, le poids, la taille, l'IMC, les antécédents familiaux, et les habitudes de vie, afin de constituer une base de données permettant d'évaluer le risque de diabète.
- **Traitement des données** : Mettre en place un algorithme de traitement des données collectées, basé sur des modèles prédictifs validés scientifiquement, pour calculer de manière précise et fiable le risque de développer le diabète.

- **Affichage des résultats de prédiction** : Créer une interface utilisateur intuitive permettant l'affichage clair et détaillé des résultats de prédiction, avec des recommandations personnalisées et la possibilité de suivre les tendances de risque au fil du temps.

4. Organisation du document

Le document est structuré essentiellement en trois parties majeures qui décrivent chacune un aspect fondamental du projet. II s'agit de :

- État de l'art : Ce premier chapitre regroupe les recherches menées sur le domaine, sur les solutions existantes ainsi que leurs éventuelles caractéristiques.
- Matériel et méthodes de conception : Le deuxième chapitre présente les outils ainsi que les techniques choisies en vue de développer le prototype de l'application.
- Résultats et discussion : Le troisième et dernier chapitre est celui où sera présenté le prototype par le biais de captures d'écrans.

Chapitre 1

État de l'art

Introduction

Le diabète est devenu l'un des plus grands défis de santé publique à l'échelle mondiale, avec une augmentation exponentielle du nombre de personnes atteintes. Selon l'Organisation mondiale de la santé (OMS), le diabète pourrait affecter plus de 700 millions de personnes d'ici 2045. La prévention et la gestion du diabète, notamment du diabète de type 2, sont devenues des priorités dans la lutte contre cette maladie. Les avancées technologiques, notamment dans le domaine de l'intelligence artificielle et des systèmes d'information de santé, ont permis de développer des outils et des applications capables d'évaluer les risques de développer le diabète et d'accompagner les patients dans leur gestion quotidienne.

Cependant, malgré la disponibilité croissante de ces applications, de nombreuses lacunes demeurent. Ces outils ne sont souvent pas adaptés aux spécificités des populations locales, notamment en Afrique, où le diabète est sous-diagnostiqué et souvent mal géré. Par ailleurs, l'accessibilité de ces solutions est souvent limitée, et elles ne permettent pas un suivi médical personnalisé, ce qui en réduit l'impact.

Dans ce chapitre, nous allons passer en revue les principales solutions existantes dans le domaine de la prédiction et de la gestion du diabète. L'analyse des avantages et des limites de ces applications servira de base pour justifier la nécessité de notre solution, *DiabOracle*, qui vise à combler ces insuffisances en proposant une approche plus adaptée et complète pour la gestion du diabète, en particulier dans le contexte africain.

1.1 Présentation des solutions existantes

Dans nos investigations, nous avons découvert cinq applications que sont : DiabTrend, One Drop, Diabet AI, MDCalc et Diabetes Australia Risk Calculator.

1.1.1 DiabTrend

DiabTrend est une application de gestion du diabète qui aide les utilisateurs à surveiller leur glycémie, à enregistrer leurs repas, et à suivre leur activité physique. Elle utilise des algorithmes d'ap-

prentissage automatique pour prédire les variations des niveaux de glucose. L'application offre des graphiques clairs pour visualiser les tendances et propose des recommandations sur les doses d'insuline à prendre.



FIGURE 1.1 : DiabTrend

1.1.2 One Drop

One Drop est une plateforme complète de gestion du diabète qui combine une application mobile avec des appareils connectés pour surveiller la glycémie. Elle permet aux utilisateurs de suivre leur alimentation, leur activité physique, et leur prise de médicaments. One Drop propose également une assistance en ligne avec des professionnels de santé certifiés, disponibles pour conseiller les utilisateurs.

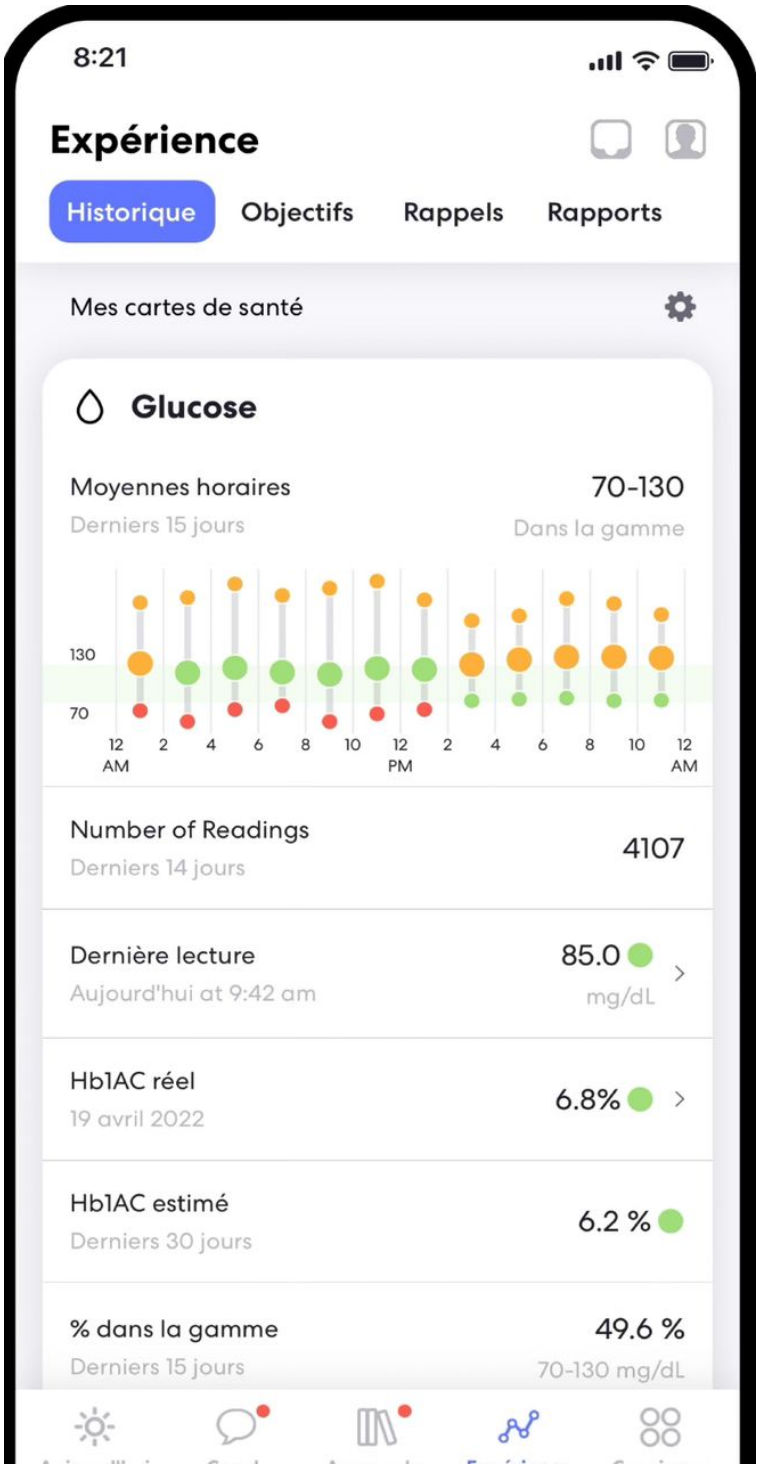


FIGURE 1.2 : Dashboard de l'application One Drop

1.1.3 Diabet AI

Diabet AI est une application innovante qui utilise l'intelligence artificielle pour prédire les risques de diabète chez ses utilisateurs. En se basant sur des algorithmes de machine learning, l'application analyse différentes données, telles que les habitudes alimentaires, les niveaux d'activité physique, les antécédents médicaux, et les taux de glucose dans le sang. Son objectif principal est de fournir une prédiction précoce du diabète de type 2 et de proposer des recommandations personnalisées pour la gestion du mode de vie.

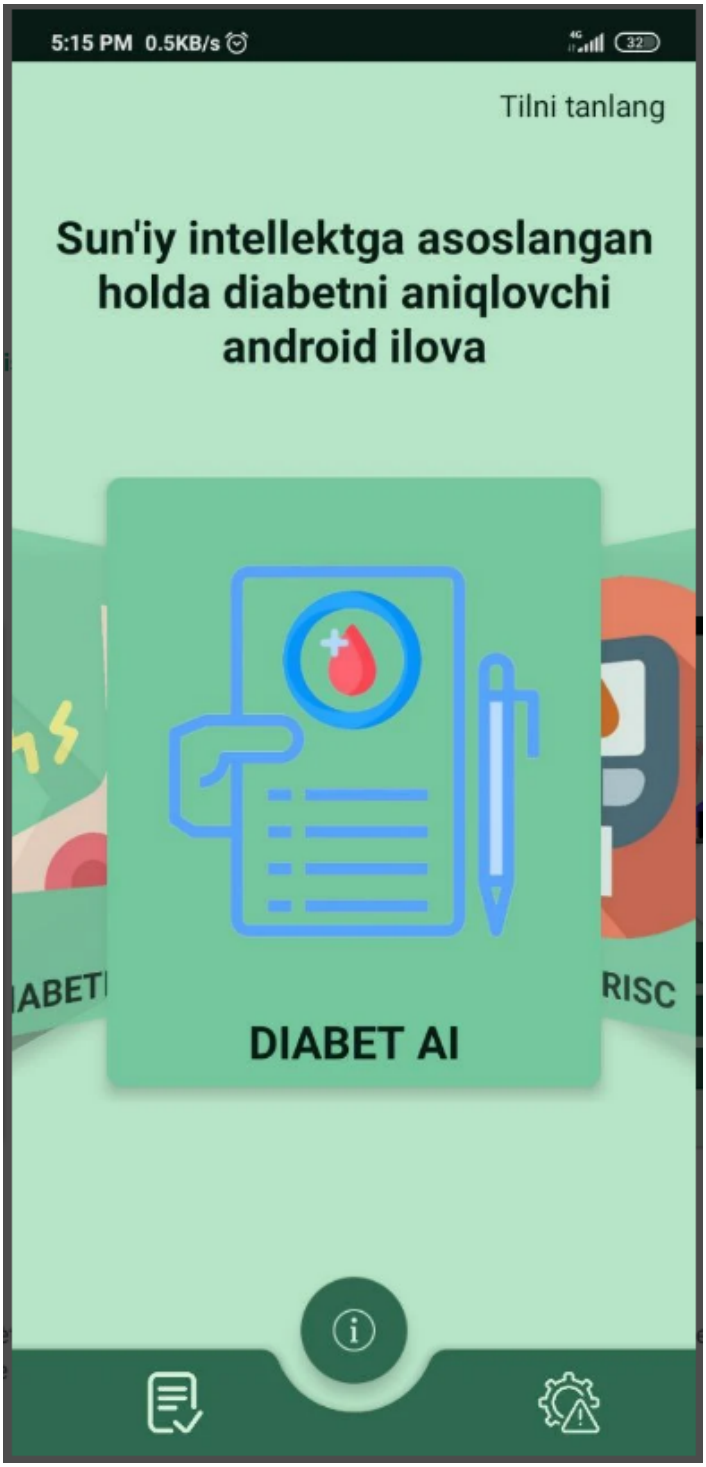


FIGURE 1.3 : Page d'acceuil de l'application Diabet AI

1.1.4 MDCalc

MDCalc est une plateforme en ligne utilisée principalement par les professionnels de santé pour accéder à une large gamme de calculatrices médicales. Ces outils aident à évaluer les risques, à diagnostiquer des conditions, et à prendre des décisions thérapeutiques basées sur des scores validés cliniquement. La calculatrice Cambridge Diabetes Risk Score (CDRS) disponible sur MDCalc est particulièrement utile pour estimer le risque de développer un diabète de type 2 chez les patients en se basant sur des données démographiques et des antécédents médicaux.

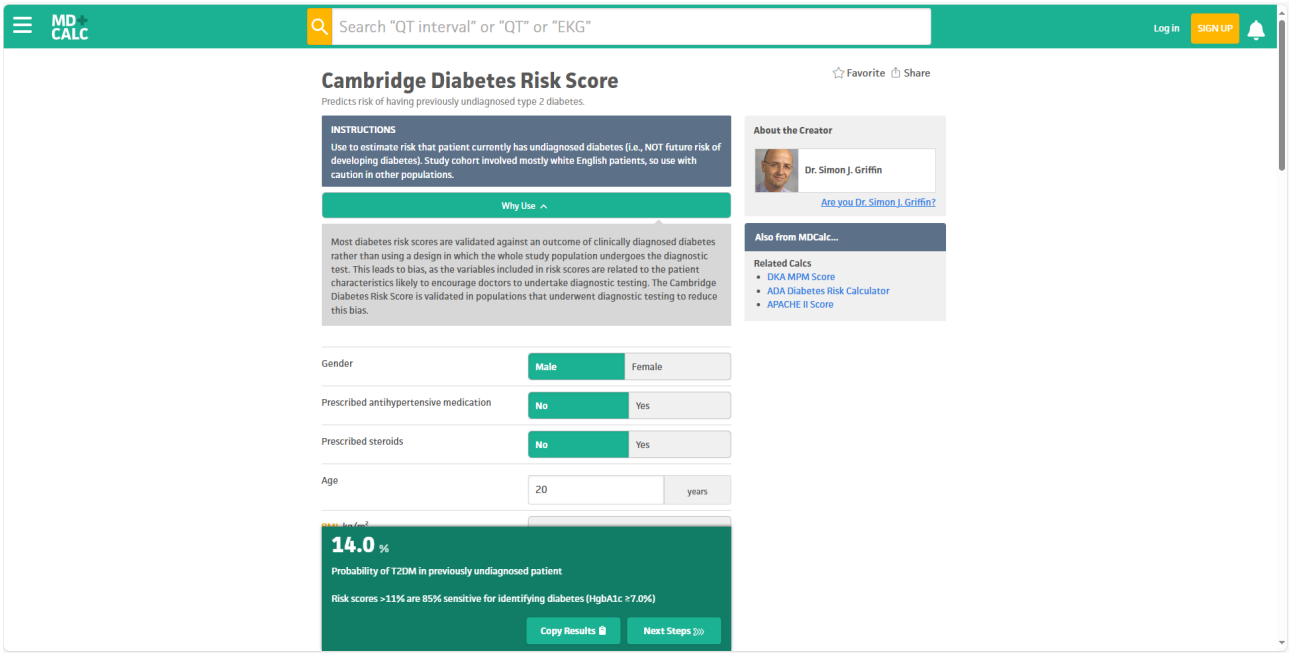


FIGURE 1.4 : Page de prédiction de l'application MD Calc

1.1.5 Diabetes Australia Risk Calculator

La Diabetes Australia Risk Calculator est un outil en ligne gratuit qui aide à évaluer le risque de développer un diabète de type 2 sur une période de cinq ans. Basée sur un questionnaire, la calculatrice prend en compte des facteurs comme l'âge, l'indice de masse corporelle (IMC), le tour de taille, les antécédents familiaux de diabète, ainsi que les habitudes de vie (alimentation, activité physique, tabagisme). Après avoir renseigné ces informations, l'utilisateur obtient un score indiquant son niveau de risque, accompagné de recommandations sur la nécessité de consulter un professionnel de santé ou d'adopter des mesures préventives.

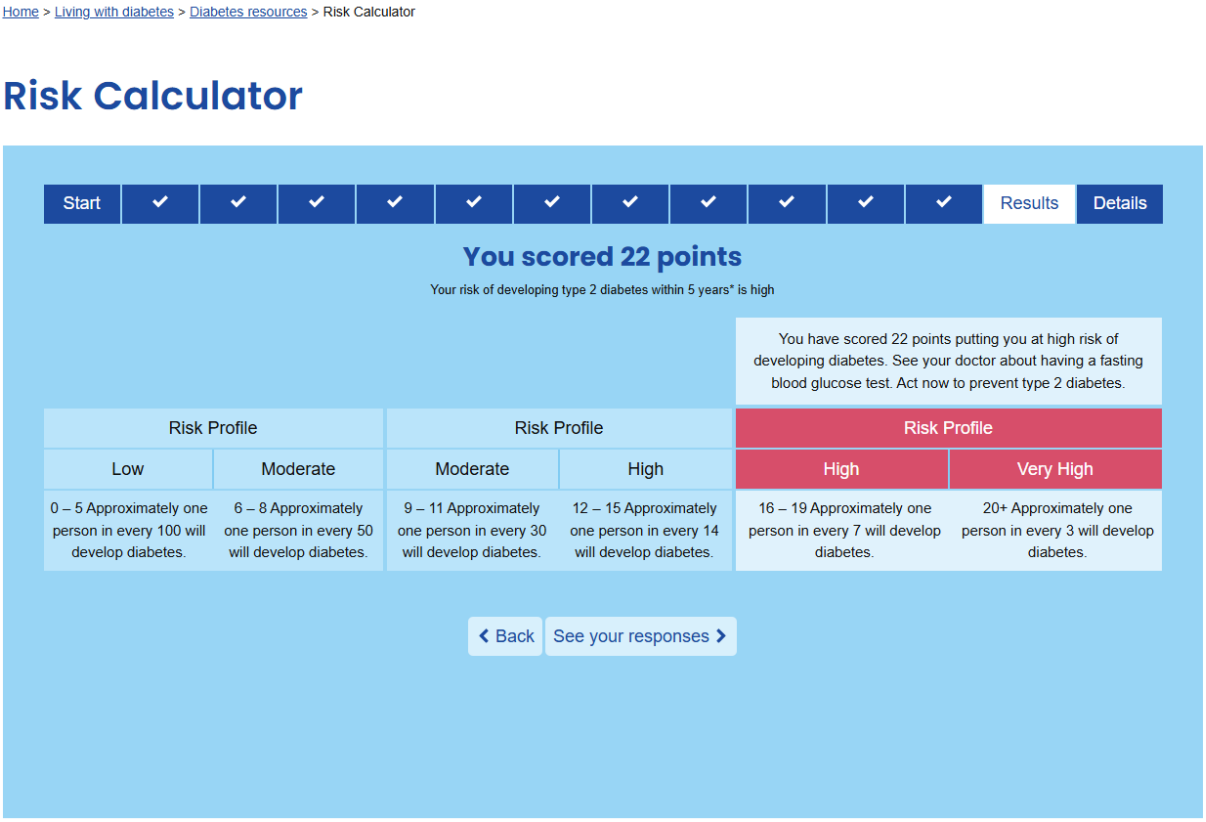


FIGURE 1.5 : Page de prédiction de l'application Diabetes Australia Risk Calculator

1.2 Comparaison des solutions existantes

Application	Avantages	Insuffisances
DiabTrend	- Suivi de la glycémie en temps réel - Prédictions de la glycémie basées sur des algorithmes d'apprentissage automatique - Graphiques clairs pour visualiser les tendances	- Ne prédit pas le diabète - Les prédictions peuvent ne pas toujours être précises pour les utilisateurs ayant des habitudes alimentaires ou des activités physiques atypiques
One Drop	- Intégration avec des appareils connectés - Accès à des professionnels de santé - Suivi complet des habitudes de vie	- Les fonctionnalités avancées et les consultations en ligne sont coûteuses ; - Dépendance à la rigueur de l'utilisateur
Diabet AI	- Utilisation de l'intelligence artificielle pour des prédictions précises - Recommandations personnalisées - Analyse de données variées	- N'est pas destinée aux professionnels de la santé, mais aux utilisateurs individuels, ce qui réduit son utilité pour le diagnostic et le suivi médical des patients diabétiques - Langage de l'application non adapté à tous les utilisateurs
MDCalc	- Large éventail de calculatrices médicales - Validation clinique des outils - Accès facile pour les professionnels de santé	- Destiné principalement aux professionnels de santé - Pas d'interaction directe avec les patients
Diabetes Australia Risk Calculator	- Outil gratuit et accessible en ligne - Sensibilisation au risque de diabète - Recommandations personnalisées en fonction du score	- Non adapté à toutes les populations - Ne prend pas en compte tous les facteurs de riqsque du diabète

TABLE 1.1 : Avantages et inconvénients des solutions de gestion du diabète existantes

Les applications de prédiction du diabète que nous avons mentionnées présentent certaines insuffisances. Tout d'abord, elles ne fournissent pas de conseils personnalisés ni de suivi médical aux utilisateurs, ce qui réduit leur impact sur la prévention et le traitement du diabète. De plus, ces applications ne sont pas destinées aux médecins, mais plutôt aux utilisateurs individuels, ce qui limite leur utilité pour le diagnostic et le suivi médical des patients diabétiques. En outre, elles ne sont pas multiplateformes, étant disponibles uniquement sur Android ou iOS, ce qui restreint leur accessibilité et leur compatibilité avec d'autres systèmes d'information.

Il est également important de noter que ces applications ne reposent pas sur des études menées en Afrique ni au Bénin, mais se basent sur des données provenant d'autres régions du monde. Cela peut affecter leur validité et leur pertinence pour la population africaine, dont la génétique peut influencer les résultats. En outre, elles ne répondent pas au besoin d'une solution de prédiction du

diabète adaptée au contexte africain, où le diabète constitue un problème de santé publique majeur, souvent négligé et sous-diagnostiqué. Enfin, ces applications ne permettent pas aux utilisateurs de suivre leurs tendances en termes de facteurs influençant le diabète (poids, taille, IMC, etc.).

Ces insuffisances mettent en lumière la nécessité d'améliorer les applications de prédiction du diabète, en utilisant des méthodes plus robustes, plus transparentes et plus respectueuses des utilisateurs.

1.3 Notre Solution

Face à ces insuffisances identifiées dans les applications de prédiction du diabète existantes, nous proposons une solution innovante : **DiabOracle**. Cette application a été spécialement conçue pour répondre aux besoins spécifiques des utilisateurs, notamment dans le contexte africain, où le diabète représente un défi de santé publique majeur.

DiabOracle intègre plusieurs fonctionnalités qui comblent les lacunes des solutions actuellement disponibles :

1. **Conseils personnalisés et suivi médical** : L'application offre des recommandations adaptées aux utilisateurs et des astuces en DIY, facile à réaliser et à moindre cout. De plus, elle permet aux utilisateurs de se connecter avec des professionnels de santé pour un suivi médical continu, garantissant ainsi un accompagnement efficace dans la gestion du diabète.

2. **Utilisation par les professionnels de santé** : DiabOracle est conçu non seulement pour les utilisateurs individuels, mais aussi pour les médecins, facilitant ainsi le diagnostic et le suivi médical des patients diabétiques. Les professionnels peuvent accéder aux données des patients et ajuster les recommandations en conséquence.

3. **Multiplateforme** : L'application sera disponible sur le web, garantissant une accessibilité optimale et une compatibilité avec divers systèmes d'information iOS, Android et Bureau. Cela permettra à un plus grand nombre d'utilisateurs d'en bénéficier.

4. **Données locales et pertinence contextuelle** : DiabOracle repose sur des études et des données spécifiques à la population africaine (Bénin), prenant en compte les facteurs génétiques et environnementaux qui influencent le diabète dans cette région. Cela garantit que les résultats et recommandations sont pertinents et valides pour les utilisateurs.

5. **Suivi des facteurs influençant le diabète** : Les utilisateurs pourront suivre leurs tendances en matière de poids, de taille, d'IMC, et d'autres facteurs de risque liés au diabète. Cette fonctionnalité permettra aux utilisateurs de mieux comprendre l'impact de leur mode de vie sur leur santé et de prendre des décisions éclairées.

En intégrant ces fonctionnalités, **DiabOracle** se positionne comme une solution complète et adaptée aux défis du diabète, offrant ainsi un outil efficace pour la prévention et la gestion de cette maladie.

Conclusion

Ce premier chapitre nous a permis d'analyser les solutions existantes dans le domaine de la prédiction du diabète, en mettant en lumière les avantages et les insuffisances de plusieurs applications, telles que *DiabTrend*, *One Drop*, *Diabet AI*, *MDCalc* et la *Diabetes Australia Risk Calculator*. Bien que ces outils offrent des fonctionnalités intéressantes, notamment en termes de gestion du diabète et de prédiction des risques, il est évident qu'ils présentent certaines limites, particulièrement dans le contexte africain. Parmi ces insuffisances, nous avons relevé l'absence de conseils personnalisés adaptés à la population africaine, le manque d'accessibilité multiplateforme et l'inadéquation avec les réalités locales, notamment en ce qui concerne les données et les études qui sous-tendent ces applications.

C'est dans ce contexte que s'inscrit le développement de notre solution, **DiabOracle**, qui a pour objectif de combler ces lacunes. En tenant compte des spécificités culturelles, environnementales et génétiques des populations africaines, cette application se propose non seulement d'améliorer la précision des prédictions, mais également d'offrir un accompagnement personnalisé et un suivi médical complet. Ainsi, *DiabOracle* constitue une alternative robuste et adaptée pour répondre aux enjeux de santé publique liés au diabète en Afrique, et plus particulièrement au Bénin. Ce travail s'inscrit donc dans une perspective d'innovation technologique et de pertinence locale, pour une meilleure prévention et gestion du diabète.

Chapitre 2

Méthodologie de travail et Conception du système

Introduction

Dans ce chapitre, nous allons détailler la méthodologie adoptée pour le développement de l'application de prédiction du diabète ainsi que la conception du système mis en œuvre. La section sur la méthodologie de travail explique les démarches systématiques utilisées pour structurer et organiser le projet, depuis l'analyse des besoins jusqu'à l'évaluation des performances. Nous aborderons les étapes clés du développement, incluant la collecte et le traitement des données, la sélection et l'évaluation des modèles de machine learning, ainsi que l'intégration de ces modèles dans l'application.

Ensuite, nous explorerons la conception du système, en mettant l'accent sur l'architecture globale de l'application, la conception de l'interface utilisateur, et les choix technologiques effectués. Nous détaillerons les choix de conception qui ont guidé le développement de l'application, les outils et technologies utilisés, ainsi que les considérations de performance et de maintenabilité.

Ce chapitre vise à fournir une compréhension approfondie des processus méthodologiques et des décisions de conception qui ont façonné le projet, permettant ainsi de saisir comment chaque étape a contribué à l'atteinte des objectifs du projet.

2.1 Analyse des Besoins

L'analyse des besoins constitue une phase cruciale dans la gestion de projet. Son objectif principal est d'identifier de manière exhaustive les exigences du projet et de dresser un état des lieux des éléments attendus. Cette étape, idéalement réalisée en amont de la conception du projet digital, offre une vision claire des attentes et des contraintes, fournissant ainsi une base solide pour la planification et l'exécution du projet. Ce chapitre se consacrera à cette analyse approfondie des besoins, jetant les bases de la réalisation d'une solution de prédiction du diabète adaptée au contexte africain en général et béninois en particulier.

2.1.1 Rappel du Contexte

Avant de plonger dans l'analyse détaillée des besoins, il est essentiel de rappeler le contexte dans lequel s'inscrit ce projet. Nous œuvrons à la création d'une application de prédiction du diabète, spécifiquement adaptée à la réalité africaine et, plus précisément, au Bénin. Les applications existantes, bien que performantes, présentent des lacunes notables, notamment dans leur incapacité à prendre en compte tous les facteurs de risque, à offrir un suivi médical personnalisé, et à s'adapter aux spécificités de la population africaine.

2.1.2 Besoins et contraintes

En vue de développer le produit le plus efficient possible, certains besoins techniques et spécifiques se sont faits ressentir.

Il s'agit de :

- L'Utilisation de la modélisation orientée objet
- L'Utilisation d'une méthode de gestion de projet adaptée aux circonstances
- L'Utilisation d'un langage de programmation orienté objet
- L'Utilisation des principes de programmation dynamique

Les besoins spécifiques sont les suivants :

- les visiteurs peuvent faire la prédiction de manière anonyme sans posséder un compte utilisateur
- il existe deux types d'utilisateurs : professionnel de la santé et visiteur

2.2 Paramètres de conception

En informatique, un paramètre représente une valeur, une option dont le choix permet d'adapter une application à l'environnement, cette section de notre mémoire présente les paramètres pris en considération pour définir les choix techniques.

La réalisation d'une application de prédiction du diabète peut être faite sous divers cadres techniques à savoir une application web, ou une application mobile ou encore un logiciel de bureau. Cela étant, le contexte particulier de ce projet nous amène à choisir un cadre précis.

L'application mobile a pour avantage d'être installable directement sur smartphone, elle est l'idéal pour une utilisation récurrente et régulière alors qu'une application web est consultable depuis un navigateur web sur tout équipement connecté à internet et disposant de son URL.

Elle est adaptée à une utilisation ciblée et limitée dans le temps. De ce fait, le choix des outils et méthodes ont convergé vers les plus performants en développement d'application web. Utiliser une application web pour notre projet de prédiction du diabète présente plusieurs avantages, tels que la réduction des coûts de développement, l'installation et la maintenance simples, l'augmentation de la productivité, et la compatibilité avec systémes d'exploitation.

2.3 Outils de conception

Le succès dans la réalisation d'un projet est dépendant de nombre de facteurs parmi lesquels figurent les outils et les méthodes appliqués pendant sa réalisation. A cet effet, nous présenterons dans cette section du document les outils qui ont été retenus pour la conception.

- Système de Contrôle de Version : Git et Github
- Langage de Programmation : HTML, CSS, JavaScript, SQL et Python
- Langage de Modélisation : UML
- Framework de développement web : Boostrap et Django
- SGBD : MySQL

2.3.1 Système de Contrôle de Version

Les logiciels de contrôle de version sont généralement utilisés en développement informatique pour assurer un meilleur suivi des projets au fil du temps. Ils gardent une trace de chaque changement apporté au code source et se révèlent très utiles en cas d'erreur grave puisqu'ils rendent possible un retour en arrière. Dans notre projet, notre choix s'est porté sur Git qui est un logiciel de gestion de versions décentralisé, créé par Linus Torvalds (Auteur du noyau de linux) et distribué selon les termes de la licence publique générale GNU version2.

Il est utilisé de façon complémentaire avec le service web d'hébergement et gestion de développement logiciel GitHub.

2.3.2 Langage de Programmation

Il s'agit d'une notation conventionnelle destinée à formuler des algorithmes et produire des programmes informatiques qui les appliquent. Les langages de programmation possèdent leur propre alphabet, vocabulaire, règles syntaxiques et sémantiques.

Dans le cadre du développement de notre application de prédiction du diabète, nous avons utilisé un ensemble diversifié de langages de programmation, chacun contribuant de manière spécifique à la réalisation du projet. **HTML (HyperText Markup Language)** a servi de base pour structurer le contenu de l'application web, en définissant les éléments de la page et leur hiérarchie. **CSS (Cascading Style Sheets)** a été utilisé pour la mise en forme et le design visuel, permettant d'améliorer l'expérience utilisateur grâce à une présentation attrayante et cohérente. **JavaScript**, quant à lui, a joué un rôle décisif dans l'interactivité de l'application, en permettant des manipulations dynamiques du contenu et des interactions en temps réel avec les utilisateurs.

Pour la gestion des données, **SQL (Structured Query Language)** a été employé pour interagir avec la base de données MySQL. Il a facilité l'exécution des requêtes nécessaires pour le stockage, la récupération et la manipulation des données relatives aux utilisateurs et à leurs résultats de prédiction. Enfin, **Python** a été le langage principal pour le développement des modèles de machine learning. Grâce à des bibliothèques telles que scikit-learn et TensorFlow, nous avons pu construire, entraîner et évaluer les algorithmes de prédiction nécessaires à notre application. L'intégration de Python avec les autres technologies a permis de créer une solution cohérente et performante, en

exploitant les forces spécifiques de chaque langage pour optimiser les différentes composantes du système.

2.3.3 Langage de Modélisation

UML (Unified Modeling Language) est un langage de modélisation graphique à base de pictogrammes conçu comme une méthode normalisée de visualisation dans les domaines du développement logiciel et en conception orientée objet. Il est destiné à faciliter la conception de documents nécessaires au développement des logiciels orienté objet par l'intermédiaire de ses diagrammes pouvant représenter :

- l'activité d'un objet
- les acteurs
- les processus
- le schéma de base de données
- composants logiciels
- réutilisation des composants

Son importance n'est donc plus à prouver pour la réalisation d'un projet informatique d'envergure.

2.3.3.1 Diagramme de cas d'utilisation

Les diagrammes de cas d'utilisation (DCU) sont des diagrammes UML utilisés pour donner une vision globale du comportement fonctionnel d'un système logiciel. Un cas d'utilisation représente une unité discrète d'interaction entre un utilisateur (humain ou machine) et un système. C'est une unité significative de travail. Dans un diagramme de cas d'utilisation, les utilisateurs sont appelés acteurs (actors), ils interagissent avec les cas d'utilisation (use cases). Ce diagramme permet d'identifier toutes les fonctionnalités que doit fournir le système ainsi que de le délimiter. Afin de le rendre plus explicite nous avons réalisé le diagramme de cas d'utilisation présenté à la figure 2.1.

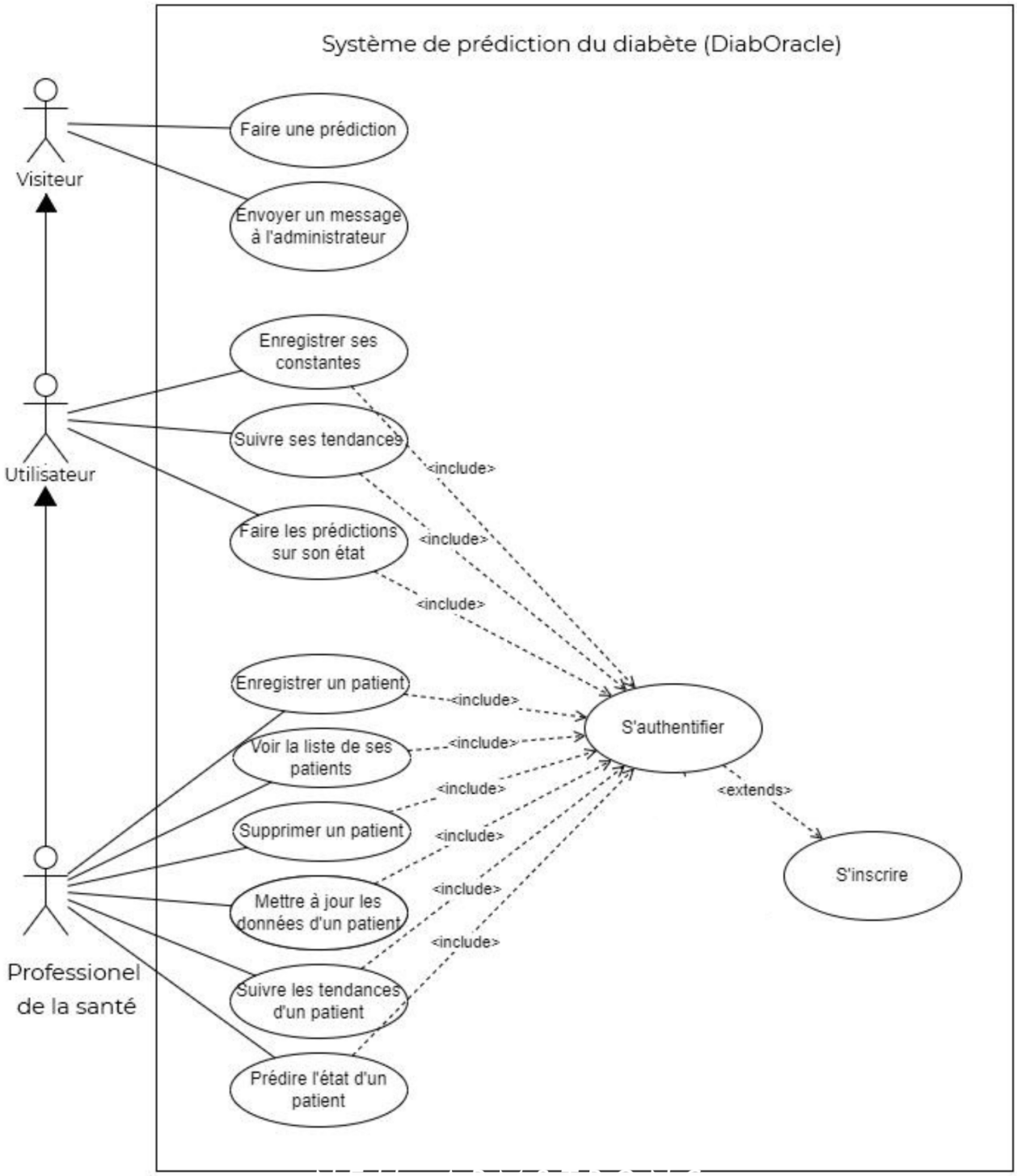


FIGURE 2.1 : Diagramme de cas d'utilisation

Le **système** présente les interactions de 3 types d'utilisateur : **un visiteur** (qui n'a pas besoin de se connecter), **un utilisateur** simple (qui gère ses propres données) et un **professionnel de la santé** (qui gère ses données puis celles de ses patients). Le visiteur peut faire des prédictions en renseignant ses données sans que ces dernières ne soient enregistrées. Il peut aussi envoyer un message à l'administrateur via un formulaire sur la page d'accueil.

L'utilisateur simple a la possibilité de saisir ses constantes journalières, de suivre l'évolution de celles-ci et d'effectuer des prédictions concernant ses données afin d'évaluer son risque de développer un diabète. De plus, il bénéficie des mêmes fonctionnalités que le visiteur.

Le professionnel de la santé, à l'instar d'un utilisateur simple, dispose de fonctionnalités similaires. Cependant, il a également la capacité d'ajouter des patients. Pour chaque patient ajouté, il peut mettre à jour les données (constantes), suivre l'évolution de ces constantes et réaliser des prédictions concernant le risque de développer un diabète. De plus, il a la possibilité de consulter la liste de ses patients et, si nécessaire, de supprimer un patient.

Il est important de noter que, pour toutes les actions spécifiques aux utilisateurs simples et aux professionnels de la santé, il est nécessaire d'être connecté. Par conséquent, une inscription préalable sur la plateforme est requise.

2.3.3.2 Diagramme de classes

Le diagramme de classes est un schéma utilisé pour présenter les classes et les interfaces des systèmes ainsi que les différentes relations entre celles-ci. Une classe décrit les responsabilités, le comportement et le type d'un ensemble d'objets. Les éléments de cet ensemble sont les instances de la classe.

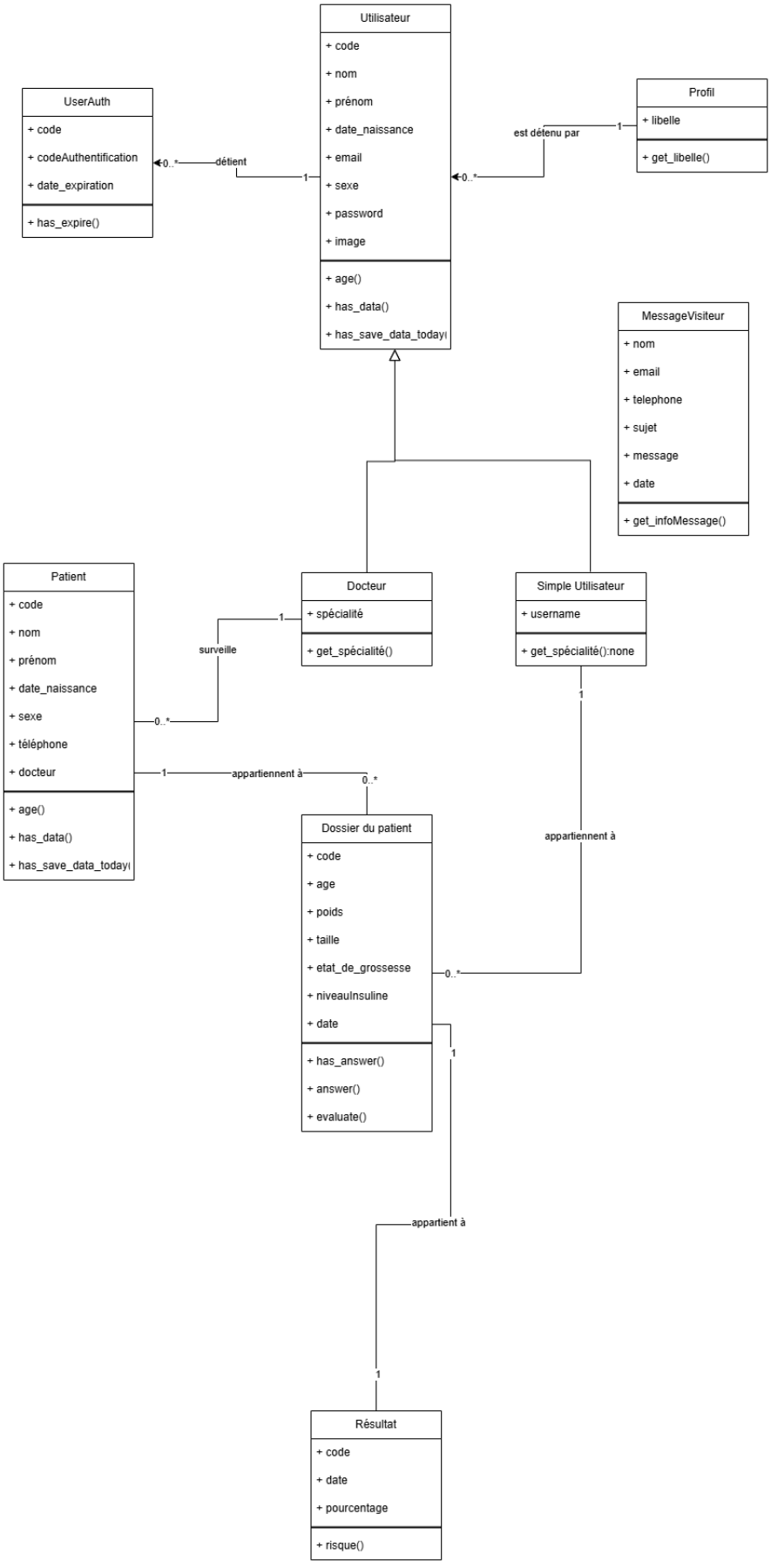


FIGURE 2.2 : Diagramme de classes

Le diagramme illustre les différentes classes du système ainsi que les interactions entre celles-ci. Un profil donné peut être associé à plusieurs utilisateurs et représente le rôle attribué à chacun d'eux. Ainsi, c'est le profil qui nous permet de déterminer si un utilisateur est un médecin ou un utilisateur simple.

Lors de la connexion, chaque utilisateur crée un objet **UserAuth**, ou authentification d'utilisateur, qui lui permet de confirmer son identité lors de la connexion à son compte. Cet objet **UserAuth** a une durée de validité de 15 minutes.

Un médecin peut surveiller aucun ou plusieurs patients. Pour un patient ou un utilisateur simple donné, il peut y avoir aucune ou plusieurs données enregistrées. Chacune de ces données représente les enregistrements des jours effectués, soit par le médecin (dans le cas d'un patient), soit par l'utilisateur simple. Chaque donnée génère un unique résultat.

Enfin, la classe **MessageVisiteur** n'interagit avec aucune autre classe, car elle ne dépend d'aucune autre classe du système.

2.3.3.3 Diagramme de séquence

Le diagramme de séquence est la représentation graphique des interactions entre les acteurs et le système selon un ordre chronologique dans la formulation UML. La figure 2.3 montre les interactions entre les acteurs et notre système.

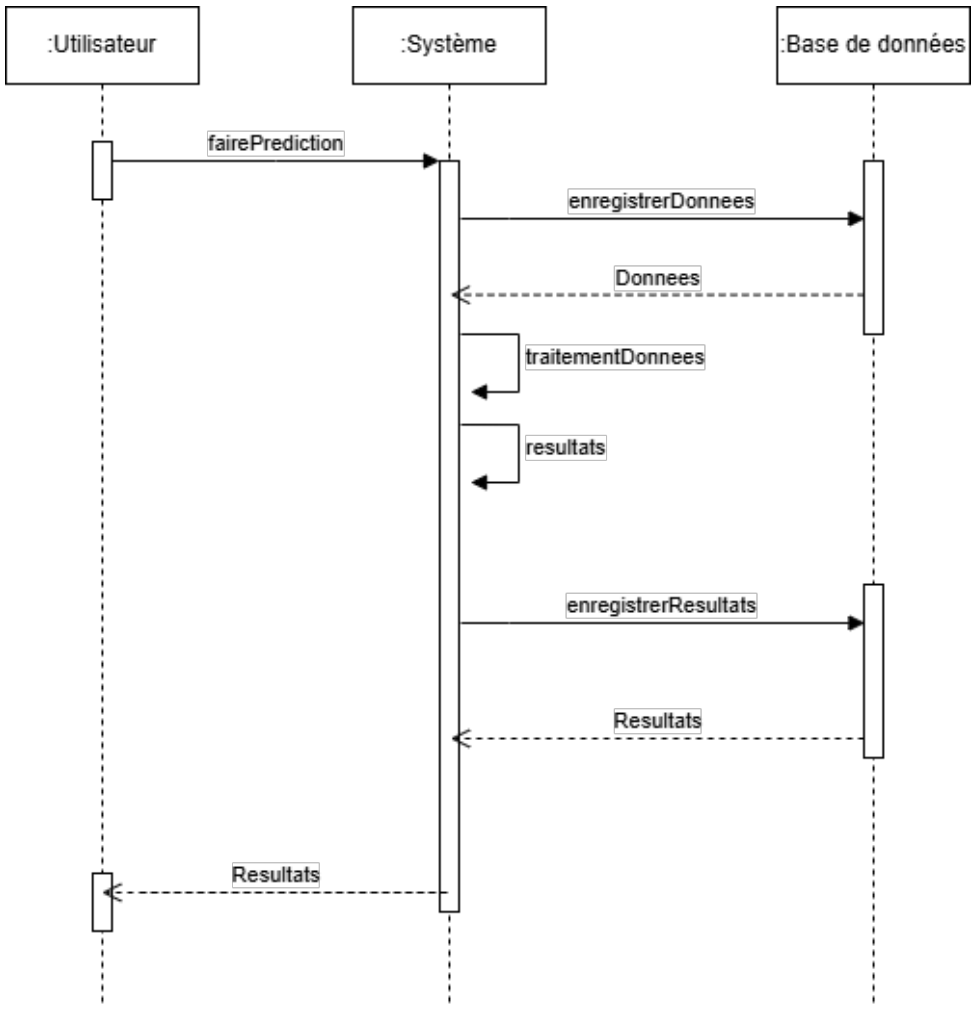


FIGURE 2.3 : Diagramme de séquences : faire une prédiction

Ce diagramme de séquence illustre l'interaction entre quatre entités : **Utilisateur**, **Système**, **Donnée**, et **Résultats**. Voici la description des étapes présentées dans le diagramme :

1. **fairePrediction** : L'**Utilisateur** initie une demande de prédiction en passant une donnée au **Système**.
2. **enregistrerDonnee** : Le **Système** enregistre la donnée en la transmettant à l'entité **Donnée** pour stockage.
3. **traitement** : Le **Système** effectue un traitement de la donnée (d) pour générer un résultat.
4. **Résultats** : Une fois le traitement terminé, le **Système** produit les **Résultats** basés sur la donnée traitée.
5. **enregistrerResultats(resultat)** : Le **Système** enregistre ensuite ces résultats en les envoyant à l'entité **Résultats**.
6. **Résultats** : Les **Résultats** sont retournés à l'**Utilisateur** comme réponse finale à sa demande initiale de prédiction.

Ce **Diagramme de séquences** montre un flux d'interaction simple où l'utilisateur fait une demande de prédiction basée sur une donnée, cette donnée est enregistrée, traitée, et les résultats sont finalement renvoyés à l'utilisateur après enregistrement. Il s'agit d'une séquence classique de gestion de prédictions dans un système de traitement de données.

2.3.4 Framework de développement web

Encore appelé infrastructure logicielle ou infrastructure de développement, c'est un ensemble cohérent de composants logiciels structurels qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel.

Pour le développement de notre application de prédiction du diabète, nous avons intégré plusieurs frameworks et outils modernes afin d'assurer une conception efficace et une expérience utilisateur optimale. Pour la conception visuelle, **Boostrap** a été employé comme framework de design utilitaire. Grâce à sa méthodologie basée sur des classes utilitaires, Boostrap a permis de créer une mise en page élégante et responsive tout en minimisant la nécessité d'écrire des CSS personnalisés complexes. Enfin, **Django**, un framework Python de haut niveau, a été utilisé pour le développement backend. Sa structure robuste et ses fonctionnalités intégrées, telles que l'ORM (Object-Relational Mapping) et le système d'authentification, ont facilité la gestion des données, la création d'API, et l'implémentation des logiques de traitement nécessaires pour le modèle de prédiction. Ensemble, ces outils ont permis d'assurer une intégration harmonieuse entre le frontend et le backend, tout en optimisant la performance et la maintenabilité de l'application.

2.3.5 Le Système de Gestion de Bases de Données

Un SGBD est un logiciel système servant à stocker, à manipuler ou gérer, et à partager des données dans une base de données en garantissant la qualité, la pérennité et la confidentialité des informations, tout en cachant la complexité des opérations. Parmi les SGBD les plus utilisés se distingue MySQL qui est un SGBDR distribué sous double licence GPL et propriétaire. MySQL fait partie du quator LAMPP(Linux Apache MySQL, PHP) très utilisé par les sites webs dynamiques.

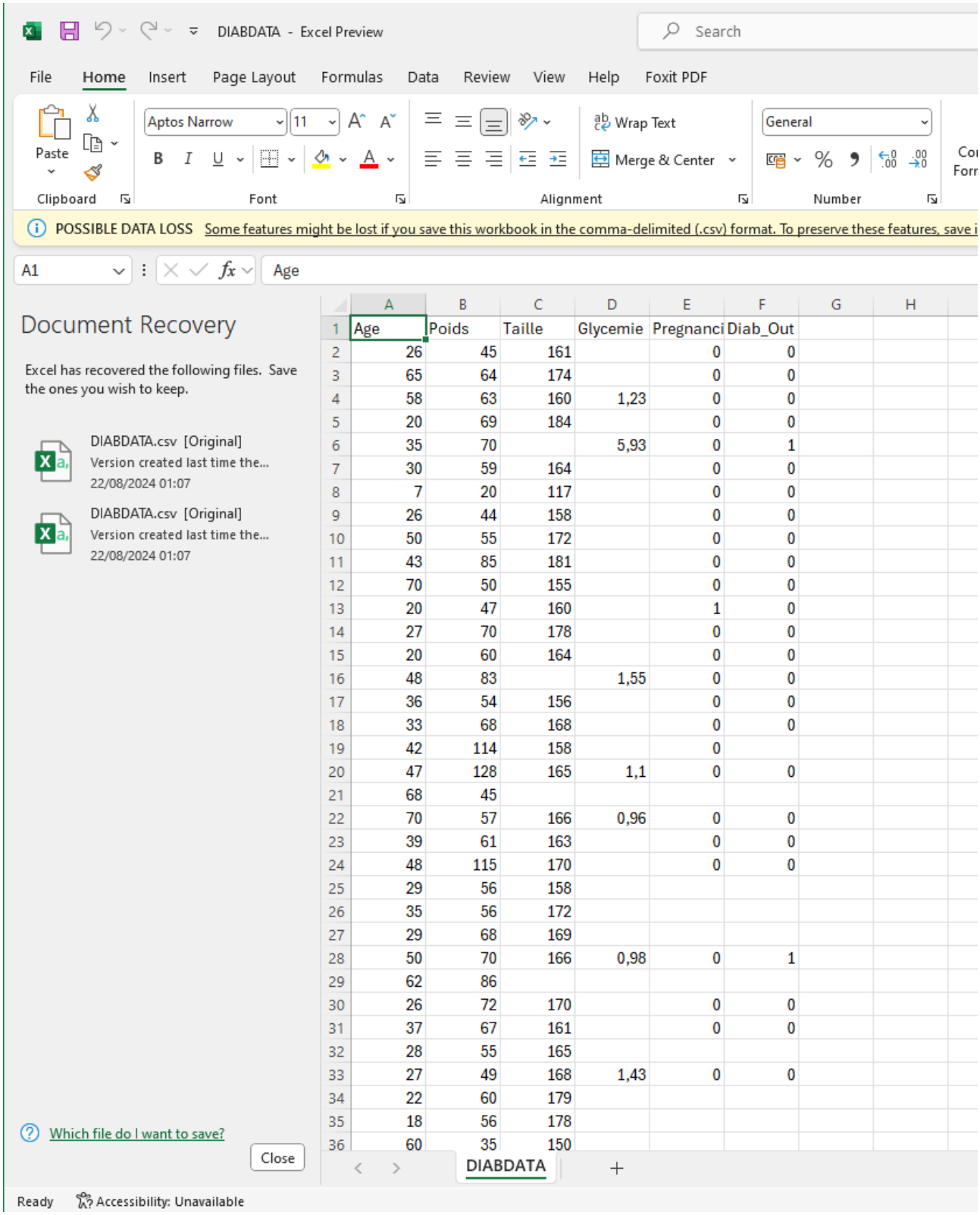
2.4 Conception et Implémentation du modèle prédictif

La conception du modèle prédictif de diabète repose sur plusieurs étapes clés, allant de la préparation des données à la sélection des algorithmes de machine learning appropriés. Notre étude se déroulant dans un contexte africain, plus précisément au Bénin, il est essentiel d'utiliser des données représentatives de la population béninoise pour garantir la pertinence et l'efficacité du modèle.

2.4.1 Besoins de données

1. **Collecte de données**

   La collecte de données pour notre étude sur la prédiction du diabète s'est déroulée du 11 septembre 2023 au 10 octobre à l'Hôpital de Zone de Lokossa. Cette collecte a été réalisée sous l'autorisation n° 1839/23/DDS-M/SPAF/SA délivrée par la Direction Départementale de la Santé du Mono. Les données ont été recueillies de manière rétrospective, en prenant en compte les patients de l'hôpital durant le premier semestre de l'année 2023.



	A	B	C	D	E	F	G	H
1	Age	Poids	Taille	Glycemie	Pregnanci	Diab_Out		
2	26	45	161		0	0		
3	65	64	174		0	0		
4	58	63	160	1,23	0	0		
5	20	69	184		0	0		
6	35	70		5,93	0	1		
7	30	59	164		0	0		
8	7	20	117		0	0		
9	26	44	158		0	0		
10	50	55	172		0	0		
11	43	85	181		0	0		
12	70	50	155		0	0		
13	20	47	160		1	0		
14	27	70	178		0	0		
15	20	60	164		0	0		
16	48	83		1,55	0	0		
17	36	54	156		0	0		
18	33	68	168		0	0		
19	42	114	158		0			
20	47	128	165	1,1	0	0		
21	68	45						
22	70	57	166	0,96	0	0		
23	39	61	163		0	0		
24	48	115	170		0	0		
25	29	56	158					
26	35	56	172					
27	29	68	169					
28	50	70	166	0,98	0	1		
29	62	86						
30	26	72	170		0	0		
31	37	67	161		0	0		
32	28	55	165					
33	27	49	168	1,43	0	0		
34	22	60	179					
35	18	56	178					
36	60	35	150					

FIGURE 2.4 : Extrait des données brutes collectées

Les informations collectées ont été anonymisées pour garantir la confidentialité des patients et ont été enregistrées dans le fichier **DIABDATA.csv**. Les paramètres suivants ont été inclus dans la collecte de données :

- **Âge** : L'âge des patients en années.
- **Poids** : Le poids des patients en kilogrammes.
- **Taille** : La taille des patients en centimètres.
- **Glycémie** : Le niveau de glycémie des patients mesuré en mg/dL.
- **Pregnancies** : Le nombre de grossesses pour les patientes féminines.
- **Diab_Out** : Indicateur de sortie de diabète (diagnostic de diabète).

Cette collecte de données nous a permis de constituer un ensemble de données représentatif de la population béninoise, fournissant ainsi une base solide pour l'entraînement et la validation de notre modèle prédictif de diabète.

2. **Préparation des données**
La préparation des données est une étape décisive dans le développement de notre modèle prédictif de diabète. Elle assure que les données utilisées sont de haute qualité, cohérentes et prêtes pour l'analyse. Voici les différentes étapes suivies pour préparer les données collectées à l'Hôpital de Zone de Lokossa :

- **Chargement des données**

  Le chargement des données est une étape fondamentale dans le processus de préparation des données pour l'analyse. Voici une explication détaillée du code utilisé pour cette étape :

  ```
  import pandas as pd
  import numpy as np
  ```

  Tout d'abord, nous importons les bibliothèques nécessaires. `pandas` est une bibliothèque puissante pour la manipulation et l'analyse des données, tandis que `numpy` est utilisée pour les opérations numériques.

  ```
  data = pd.read_csv('DIABDATA.csv', sep=';')
  data
  ```

  Ensuite, nous utilisons la fonction `pd.read csv` de la bibliothèque `pandas` pour lire le fichier CSV contenant les données, nommé `DIABDATA.csv`Le paramètre `sep=';'` indique que les valeurs dans le fichier CSV sont séparées par des points-virgules. Cette ligne de code charge les données dans un DataFrame `pandas` appelé `data` qui est ensuite affiché pour vérification.

  ```
  cols_to_drop = data.filter(like='Unnamed').columns
  data = data.drop(cols_to_drop, axis=1)
  data
  ```

  Dans cette partie du code, nous nettoyons les données en supprimant les colonnes non désirées. La méthode `filter` avec l'argument `like='Unnamed'` permet de sélectionner

toutes les colonnes dont le nom contient 'Unnamed'. Ces colonnes sont souvent générées par défaut lors de la lecture de fichiers CSV et peuvent contenir des données inutiles ou des index. Les colonnes sélectionnées sont stockées dans la variable `cols to drop`Ènsuite, la méthode `drop` est utilisée pour supprimer ces colonnes du DataFrame `data` en spécifiant `axis=1` pour indiquer que nous supprimons des colonnes (et non des lignes). Le DataFrame nettoyé est ensuite affiché pour vérification.

dataCleaning.ipynb | DIABDATA.csv | Code | Share | anaconda-panel-2023.05-py310

```
[18]: import pandas as pd
      import numpy as np
```

```
[20]: data = pd.read_csv('DIABDATA.csv', sep=';')
      data
```

[20]:

	Age	Poids	Taille	Glycemie	Pregnancies	Diab_Out	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 229	Unnamed: 230	Unnamed: 231
0	26.0	45.0	161	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
1	65.0	64.0	174	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2	58.0	63.0	160	1,23	0.0	0.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
3	20.0	69.0	184	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
4	35.0	70.0	NaN	5,93	0.0	1.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
737	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
738	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
739	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
740	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
741	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

742 rows × 239 columns

```
[26]: cols_to_drop=data.filter(like='Unnamed').columns
      data=data.drop(cols_to_drop, axis=1)
      data
```

[26]:

	Age	Poids	Taille	Glycemie	Pregnancies	Diab_Out
0	26.0	45.0	161	NaN	0.0	0.0
1	65.0	64.0	174	NaN	0.0	0.0
2	58.0	63.0	160	1,23	0.0	0.0
3	20.0	69.0	184	NaN	0.0	0.0

FIGURE 2.5 : Chargement des données et resultats

- **Nettoyage des données**

Le nettoyage des données est une étape essentielle pour garantir la qualité et la fiabilité des analyses ultérieures. Cette étape consiste à identifier et corriger les erreurs, les incohérences et les valeurs manquantes dans l'ensemble de données. Dans notre étude, plusieurs actions ont été entreprises pour nettoyer les données collectées à l'Hôpital de Zone de Lokossa. Tout d'abord, les enregistrements contenant des valeurs manquantes pour des variables critiques telles que l'âge, le poids, la taille et la glycémie ont été soit supprimés, soit imputés en utilisant des méthodes statistiques appropriées. Par la suite, les colonnes non désirées, souvent générées par défaut lors de la lecture de fichiers CSV, ont été supprimées pour assurer que seules les données pertinentes sont utilisées dans les analyses.

Ce processus de nettoyage rigoureux permet de travailler avec un ensemble de données propre et structuré, prêt pour les étapes ultérieures de modélisation et d'analyse.

dataCleaning.ipynb DIABDATA.csv

Code

```
[12]: data.describe(include='all')
```

[12]:

	Age	Poids	Taille	Glycemie	Pregnancies	Diab_Out
count	378.000000	357.000000	318	282	296.000000	188.000000
unique	NaN	NaN	48	126	NaN	NaN
top	NaN	NaN	168	0,93	NaN	NaN
freq	NaN	NaN	22	11	NaN	NaN
mean	45.148148	65.957983	NaN	NaN	0.013514	0.196809
std	16.982571	16.923520	NaN	NaN	0.115655	0.398648
min	7.000000	20.000000	NaN	NaN	0.000000	0.000000
25%	31.000000	55.000000	NaN	NaN	0.000000	0.000000
50%	45.000000	64.000000	NaN	NaN	0.000000	0.000000
75%	57.000000	75.000000	NaN	NaN	0.000000	0.000000
max	98.000000	133.000000	NaN	NaN	1.000000	1.000000

FIGURE 2.6 : Description des données avant nettoyage

```
data = data.dropna(subset=['Glycemie '])
```

Cette ligne de code supprime toutes les lignes du DataFrame `data` où la colonne `Glycemie` contient des valeurs manquantes (NaN). Cela garantit que seules les lignes avec des valeurs de glycémie valides sont conservées.

```
data['Pregnancies '] = data['Pregnancies '].fillna(0)
```

Ici, les valeurs manquantes dans la colonne `Pregnancies` sont remplacées par 0. Cela est pertinent car l'absence de données pour cette colonne signifie qu'il n'y a eu aucune grossesse.

```
data['Poids'] = data['Poids'].fillna(data['Poids'].mean())
```

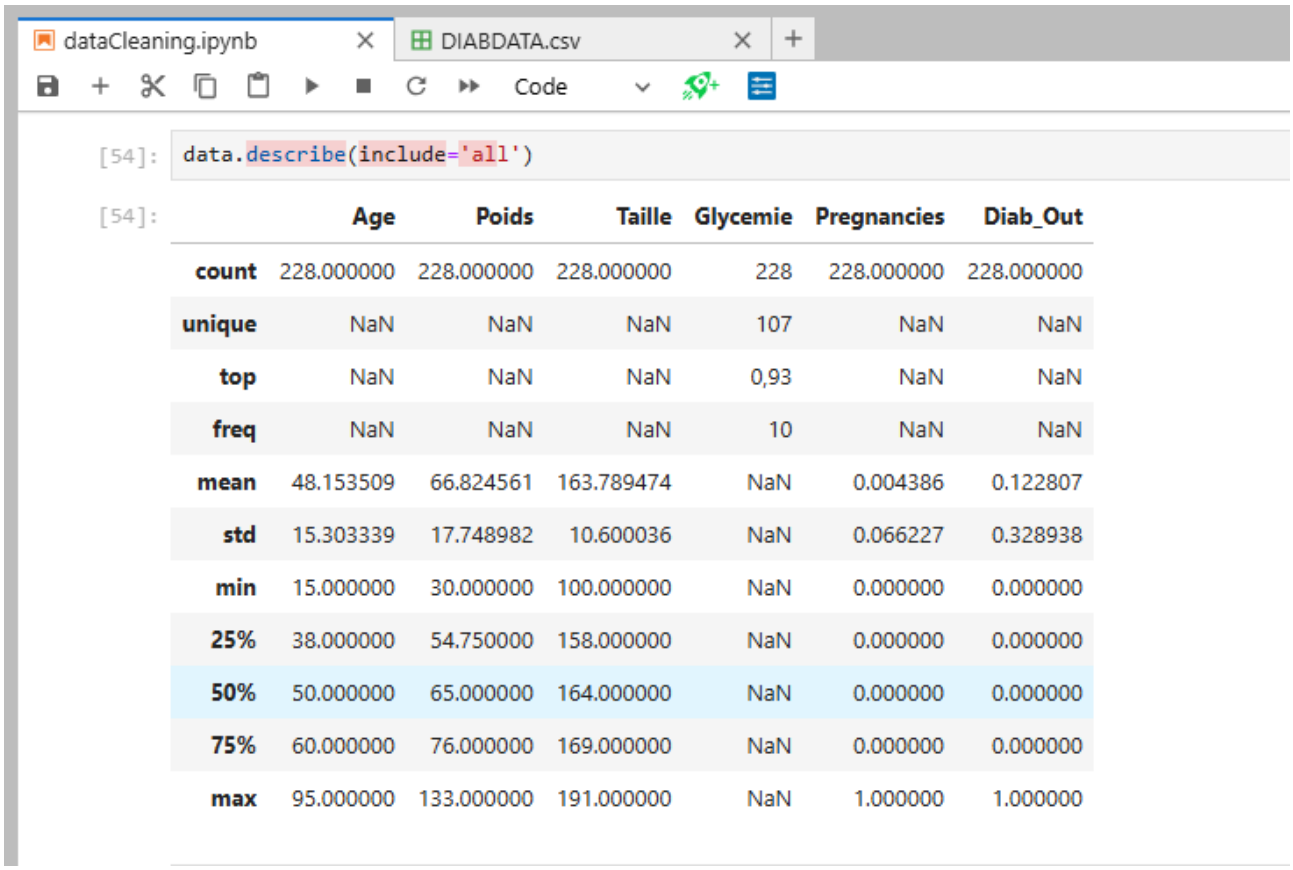
Cette ligne remplace les valeurs manquantes dans la colonne `Poids` par la moyenne des poids présents dans le DataFrame. Cela permet de conserver les lignes tout en imputant une valeur raisonnable pour les poids manquants.

```
data = data.dropna(subset=['Taille'])
```

Cette ligne supprime toutes les lignes où la colonne `Taille` contient des valeurs manquantes. Cela garantit que seules les lignes avec des valeurs de taille valides sont conservées.

```
data['Diab_Out'] = data['Diab_Out'].fillna(0)
```

Enfin, cette ligne remplace les valeurs manquantes dans la colonne `Diab_Out` par 0. Ceci signifie qu'il n'y a pas de diagnostic de diabète pour ces patients.



	Age	Poids	Taille	Glycemie	Pregnancies	Diab_Out
count	228.000000	228.000000	228.000000	228	228.000000	228.000000
unique	NaN	NaN	NaN	107	NaN	NaN
top	NaN	NaN	NaN	0,93	NaN	NaN
freq	NaN	NaN	NaN	10	NaN	NaN
mean	48.153509	66.824561	163.789474	NaN	0.004386	0.122807
std	15.303339	17.748982	10.600036	NaN	0.066227	0.328938
min	15.000000	30.000000	100.000000	NaN	0.000000	0.000000
25%	38.000000	54.750000	158.000000	NaN	0.000000	0.000000
50%	50.000000	65.000000	164.000000	NaN	0.000000	0.000000
75%	60.000000	76.000000	169.000000	NaN	0.000000	0.000000
max	95.000000	133.000000	191.000000	NaN	1.000000	1.000000

FIGURE 2.7 : Description des données après nettoyage

Ce processus de chargement et de nettoyage des données garantit que nous travaillons avec un ensemble de données propre et structuré, prêt pour les étapes ultérieures d'analyse et de modélisation.

3. **Entrainement du modèle**

Dans le cadre de notre étude, nous cherchons à prédire la probabilité qu'un individu soit atteint de diabète en fonction de plusieurs caractéristiques cliniques. Cette tâche constitue un problème de classification binaire, où l'objectif est de classer les individus en deux catégories : ceux qui sont atteints de diabète et ceux qui ne le sont pas. Cependant, notre but ne se limite pas à cette classification binaire. Nous visons également à déterminer le risque imminent que les individus développent la maladie. Pour résoudre ce problème, nous avons choisi d'utiliser la régression logistique.

La régression logistique est particulièrement adaptée pour ce type de problème de classification binaire pour plusieurs raisons. Tout d'abord, elle offre une grande interprétabilité, car elle

fournit des coefficients qui peuvent être analysés pour comprendre l'impact de chaque caractéristique sur la probabilité d'être atteint de diabète. De plus, elle permet de prédire des probabilités, ce qui est utile pour évaluer le risque de diabète et prendre des décisions médicales basées sur ces probabilités. En termes d'efficacité, la régression logistique est un modèle simple et performant qui fonctionne bien même avec des ensembles de données de taille modérée. Enfin, elle est robuste et moins sujette au surapprentissage par rapport à des modèles plus complexes, surtout lorsque le nombre de caractéristiques est limité.

dataCleaning.ipynb | Cleaner.ipynb | DIABDATA.csv

Code

[55]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

[60]:
```
# Remplacer les virgules par des points et convertir en float
data = data.replace(',', '.', regex=True)  # Remplace les virgules par des points
data = data.astype(float)  # Convertit les colonnes en float
```

[63]:
```
# Calculer la corrélation
correlation = data.corr()
print(correlation)
```

```
                Age     Poids   Taille   Glycemie  Pregnancies  Diab_Out
Age        1.000000  0.012826 -0.084482  0.066649     0.029253  0.152069
Poids      0.012826  1.000000  0.011706  0.186217     0.137773  0.116467
Taille    -0.084482  0.011706  1.000000  0.074622    -0.494274  0.030824
Glycemie   0.066649  0.186217  0.074622  1.000000    -0.035790  0.322801
Pregnancies 0.029253 0.137773 -0.494274 -0.035790     1.000000  0.108228
Diab_Out   0.152069  0.116467  0.030824  0.322801     0.108228  1.000000
```

[139]:
```
X=data.drop("Diab_Out", axis=1)
Y=data['Diab_Out']

X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size=0.2)
```

[72]:
```
model = LogisticRegression()
model.fit(X_train, Y_train)
```

[72]:
```
▾ LogisticRegression
LogisticRegression()
```

[75]:
```
predictions = model.predict(X_test)
predictions
```

[75]:
```
array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
       0., 1., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])
```

[78]:
```
accuracy =accuracy_score(predictions, Y_test)
print(accuracy)
```

FIGURE 2.8 : Entrainement du modèle

2.4.2 Mesures de performances du modèle

Pour évaluer la performance de notre modèle de prédiction du diabète, nous avions utilisé une métrique clé. Cette mesure nous permet de comprendre la précision et l'efficacité du modèle, ainsi que d'identifier les domaines nécessitant des améliorations.

Exactitude (Accuracy)

L'exactitude est la proportion des prédictions correctes parmi l'ensemble des prédictions. Elle est calculée comme suit :

$$\text{Exactitude} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}} \tag{2.1}$$

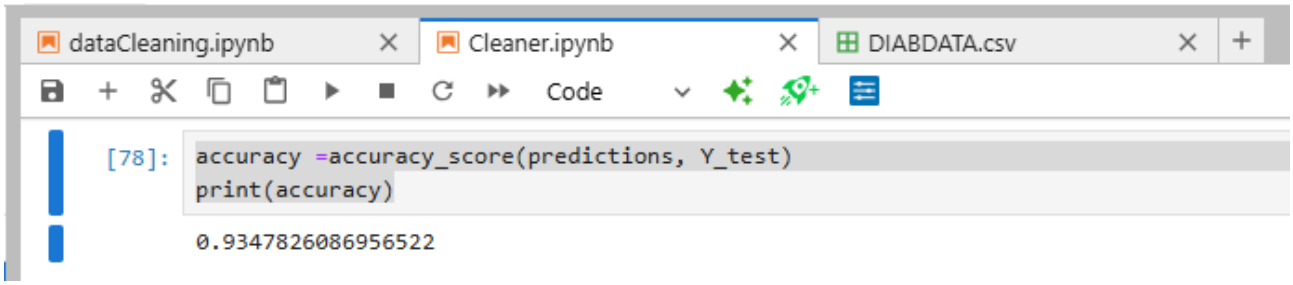


FIGURE 2.9 : Score de précision du modèle

Le résultat de 0.9347826086956522 signifie que le modèle a une exactitude de 93.48%. Cela indique que le modèle a correctement prédit environ 93.48% des échantillons de l'ensemble de test.

2.4.3 Evaluation du risque P

Pour calculer le risque de diabète en fonction des paramètres que nous avions collectés (âge, poids, taille, glycémie, état de grossesse), nous utilsons la formule de la régression logistique. Voici une formule simplifiée pour illustrer comment cela pourrait fonctionner :

Formule de Régression Logistique

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}} \tag{2.2}$$

Où :

- $P(Y = 1)$ est la probabilité que le patient soit diabétique.
- β_0 est l'ordonnée à l'origine (intercept).
- $\beta_1, \beta_2, \ldots, \beta_n$ sont les coefficients des variables indépendantes.
- $X_1, X_2, \ldots, X_n$ sont les variables indépendantes (âge, poids, taille, glycémie, état de grossesse).

Exemple de Formule

Pour nos paramètres spécifiques, la formule pourrait ressembler à ceci :

$$P(\text{diabète}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot \text{âge} + \beta_2 \cdot \text{poids} + \beta_3 \cdot \text{taille} + \beta_4 \cdot \text{glycémie} + \beta_5 \cdot \text{état de grossesse})}} \tag{2.3}$$

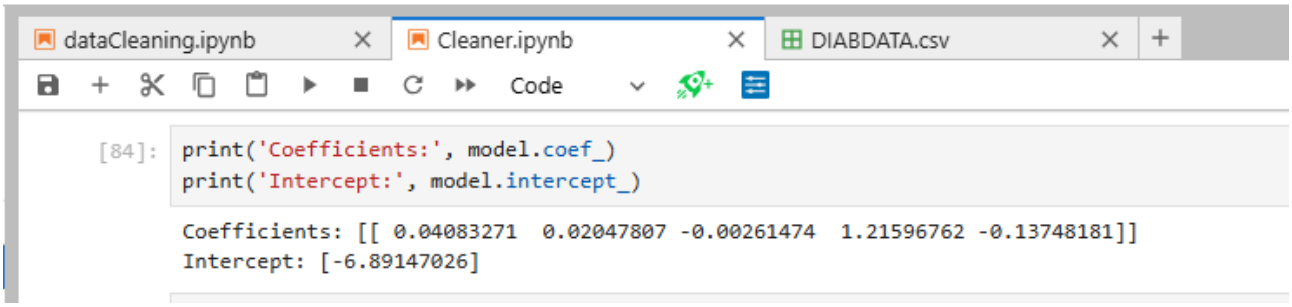


FIGURE 2.10 : Constantes du risque P

En utilisant ces coefficients et l'intercept, la formule de régression logistique pour prédire la probabilité de diabète ($P(\text{diabète})$) devient en substituant les valeurs des coefficients et de l'intercept :

$$P(\text{diabète}) = \frac{1}{1 + e^{-(-6.89 + 0.040 \cdot \text{âge} + 0.020 \cdot \text{poids} - 0.0026 \cdot \text{taille} + 1.216 \cdot \text{glycémie} - 0.137 \cdot \text{état de grossesse})}}$$

```python
import numpy as np

def predict_diabetes(age, poids, taille, glycemie,
etat_de_grossesse):
    # Coefficients obtenus du modèle
    beta_0 = -6.89147026
    beta_1 = 0.04083271
    beta_2 = 0.02047807
    beta_3 = -0.00261474
    beta_4 = 1.21596762
    beta_5 = -0.13748181

    # Calcul de la somme pondérée des paramètres
    linear_combination = (beta_0 +
                          beta_1 * age +
                          beta_2 * poids +
                          beta_3 * taille +
                          beta_4 * glycemie +
                          beta_5 * etat_de_grossesse)

    # Calcul de la probabilité en utilisant la fonction sigmoïde
    probability = 1 / (1 + np.exp(-linear_combination))

    return probability

# Exemple d'utilisation
age = 50
poids = 70
taille = 170
glycemie = 5.5
etat_de_grossesse = 0  # 0 pour non enceinte, 1 pour enceinte

probabilite_diabete = predict_diabetes(age, poids, taille,
glycemie, etat_de_grossesse)
print(f'Probabilité de diabète: {probabilite_diabete:.2f}')
```

FIGURE 2.11 : Fonction pour impléménter le risque P

2.5 Sécurité

La sécurité est un enjeu majeur dans le développement de toute application, surtout lorsqu'il s'agit de données médicales et personnelles. Dans le cadre de notre application de prédiction du diabète, plusieurs aspects de sécurité ont été pris en compte pour assurer la confidentialité, l'intégrité et la disponibilité des informations traitées.

2.5.1 Confidentialité des données

La confidentialité des données est essentielle pour protéger les informations personnelles des utilisateurs. Les données médicales étant particulièrement sensibles, il est indispensable de garantir que celles-ci ne soient accessibles qu'aux personnes autorisées. Pour cela, notre application intègre plusieurs mécanismes de sécurité :

- **Cryptage des données** : Toutes les données personnelles et médicales collectées via l'application sont cryptées à l'aide d'algorithmes de cryptographie robustes. En particulier, nous utilisons l'algorithme de Vigenère pour crypter les mots de passe stockés dans la base de données, garantissant ainsi une sécurité renforcée durant le stockage et le transfert de ces informations sensibles.

  Algorithme de Vigenère

  L'algorithme de Vigenère est un chiffrement par substitution polyalphabétique. Voici les étapes de l'algorithme :

Algorithme 1 : Vigénère

- **Initialisation** : Soit P le texte en clair de longueur n, K la clé de longueur m.
- **Construction de la clé étendue** : Répéter la clé K jusqu'à obtenir une clé étendue K' de longueur n.
- **Chiffrement** : Pour chaque caractère P_i du texte en clair et K'_i de la clé étendue :

$$C_i = (P_i + K'_i) \mod 26$$

  où C_i est le caractère chiffré correspondant.
- **Texte chiffré** : Le texte chiffré C est la concaténation de tous les C_i.

  Exemple

  Supposons que le texte en clair soit `HELLO` et la clé soit `KEY`.

  1. **Initialisation** : $P =$ `HELLO`, $K =$ `KEY`.
  2. **Construction de la clé étendue** : $K' =$ `KEYKE`.

3. **Chiffrement** :

$$
\begin{aligned}
H + K &= 7 + 10 = 17 \quad (\texttt{R}) \\
E + E &= 4 + 4 = 8 \quad (\texttt{I}) \\
L + Y &= 11 + 24 = 35 \mod 26 = 9 \quad (\texttt{J}) \\
L + K &= 11 + 10 = 21 \quad (\texttt{V}) \\
O + E &= 14 + 4 = 18 \quad (\texttt{S})
\end{aligned}
$$

4. **Texte chiffré** : $C = \texttt{RIJVS}$.

- **Anonymisation des données** : Afin de respecter les lois et régulations sur la protection des données, nous avons mis en place des techniques d'anonymisation des données. Ces techniques garantissent que les informations ne peuvent être retracées jusqu'à un individu sans une autorisation spécifique.
- **Authentification forte** : L'accès aux comptes des utilisateurs se fait par un processus sécurisé impliquant l'utilisation de mots de passe robustes et, si nécessaire, une authentification à deux facteurs (2FA). Ce mécanisme réduit considérablement les risques d'accès non autorisé.

2.5.2 Intégrité des données

L'intégrité des données vise à garantir que les informations stockées ou transmises ne soient pas altérées ou corrompues de manière malveillante ou accidentelle. Pour cela, plusieurs stratégies ont été mises en place :

- **Contrôle de version des données** : Chaque modification des données personnelles ou médicales des utilisateurs est tracée grâce à un système de contrôle de version. Cela permet de retracer les modifications et de revenir à des versions antérieures en cas d'erreurs ou d'attaques.
- **Vérification des données saisies** : Des mécanismes de validation des données sont intégrés dans l'application pour s'assurer que les informations saisies par les utilisateurs (niveaux de glycémie, poids, etc.) soient cohérentes et dans des plages valides.

2.5.3 Protection contre les menaces externes

L'application est également protégée contre les attaques informatiques qui ciblent souvent les systèmes manipulant des données sensibles. Parmi les mesures prises, on retrouve :

- **Protection contre les attaques par force brute et injection SQL** : Des systèmes sont en place pour empêcher les tentatives de connexion répétées par force brute ainsi que les attaques par injection SQL ou XSS.

2.5.4 Conformité légale

En plus des mesures techniques, il est essentiel que l'application respecte les lois et régulations en matière de protection des données personnelles, notamment les directives locales et internationales :

- **Conformité aux normes de l'APDP** : l'application respecte les directives imposées par l'APDP, garantissant le droit à l'oubli, la portabilité des données et la transparence quant à leur traitement.
- **Consentement éclairé** : Avant la collecte de toute information personnelle, les utilisateurs doivent explicitement consentir au traitement de leurs données. Ils sont informés de la manière dont ces données seront utilisées, stockées et protégées.

Conclusion

Ce chapitre consacré à l'analyse des besoins a jetté les bases nécessaires et la conception et l'implémentation d'une solution de prédiction du diabète adaptée à la réalité africaine, visant à surmonter les insuffisances des applications existantes. La compréhension approfondie des besoins constitue le socle sur lequel reposera le succès de notre projet.

Chapitre 3

Résultats et discussions

Introduction

Le diabète, une maladie chronique touchant des millions de personnes à travers le monde, nécessite une prise en charge précoce et efficace. Dans cette optique, la conception d'une application de prédiction du diabète revêt une importance capitale. Ce chapitre se penche sur le prototype de cette application novatrice, qui vise à anticiper le risque de diabète chez les individus. Avant d'entrer dans les détails techniques, il est essentiel de contextualiser notre démarche. Notre objectif est double : d'une part, créer un outil pratique pour les utilisateurs soucieux de leur santé, et d'autre part, explorer les possibilités offertes par l'apprentissage automatique dans le domaine médical.

3.1 Présentation du prototype

Le prototype de notre application combine l'expertise en développement web et en analyse de données. Il offre une interface intuitive permettant aux utilisateurs de saisir leurs informations médicales, telles que les antécédents familiaux, les habitudes alimentaires et les niveaux d'activité physique. Grâce à des algorithmes sophistiqués, l'application évalue le risque individuel de développer un diabète de type 2 .

Description du Prototype :

Le prototype de notre solution comprend une application web intégrant notre modèle prédictif. Les figures ci-après illustrent le processus de prédiction du diabète. Une particularité de notre application est que les visiteurs peuvent effectuer des prédictions de manière anonyme, sans que leurs informations soient enregistrées dans la base de données. Tout d'abord, le visiteur saisit ses informations personnelles et physiques ainsi que son taux d'insuline pour obtenir son pourcentage de risque.



FIGURE 3.1 : Page D'accueil

Ceci est notre page d'accueil, qui est subdivisée en quatre sections :

1. **Prédiction** : Dans cette première section, vous pouvez faire une prédiction de manière anonyme ou vous connecter pour suivre vos tendances.
2. **À propos** : Cette section fournit des informations sur l'application DiabOracle et son fonctionnement.
3. **Nos témoignages** : Vous pouvez lire les témoignages d'utilisateurs qui partagent leur expérience avec l'application.
4. **Contactez-nous** : Cette section vous permet de contacter l'administrateur via un formulaire.

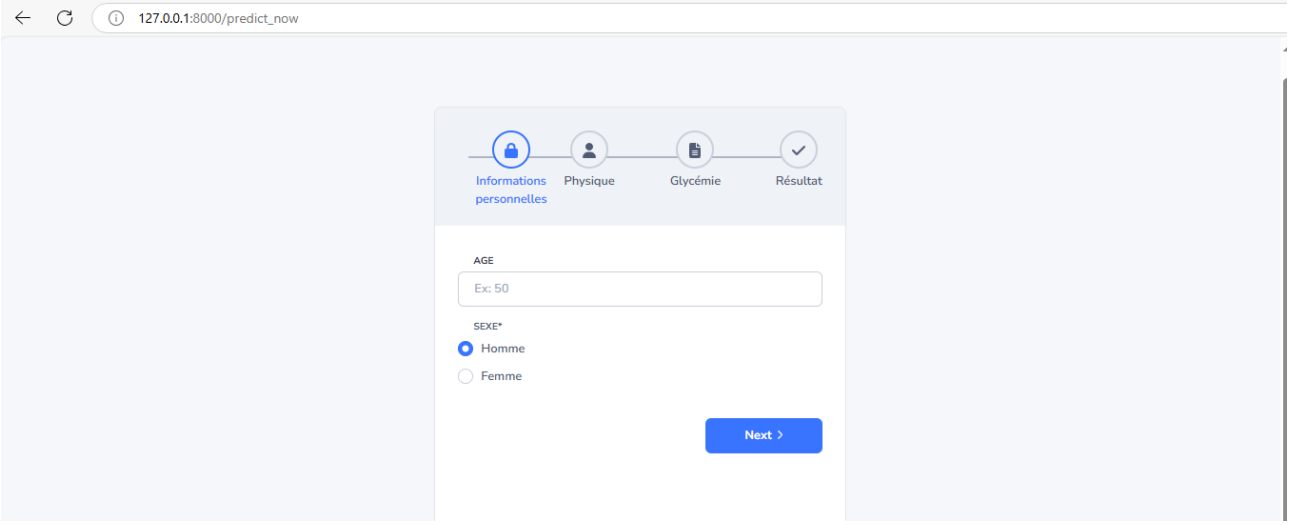


FIGURE 3.2 : Informations personnelles

Ici, nous sommes à la phase de collecte de données pour la prédiction. Cela se fait via un formulaire où l'utilisateur fournit ses informations personnelles, physiques, ainsi que sa glycémie, afin d'obtenir un résultat qui représente le risque encouru d'être atteint du diabète.

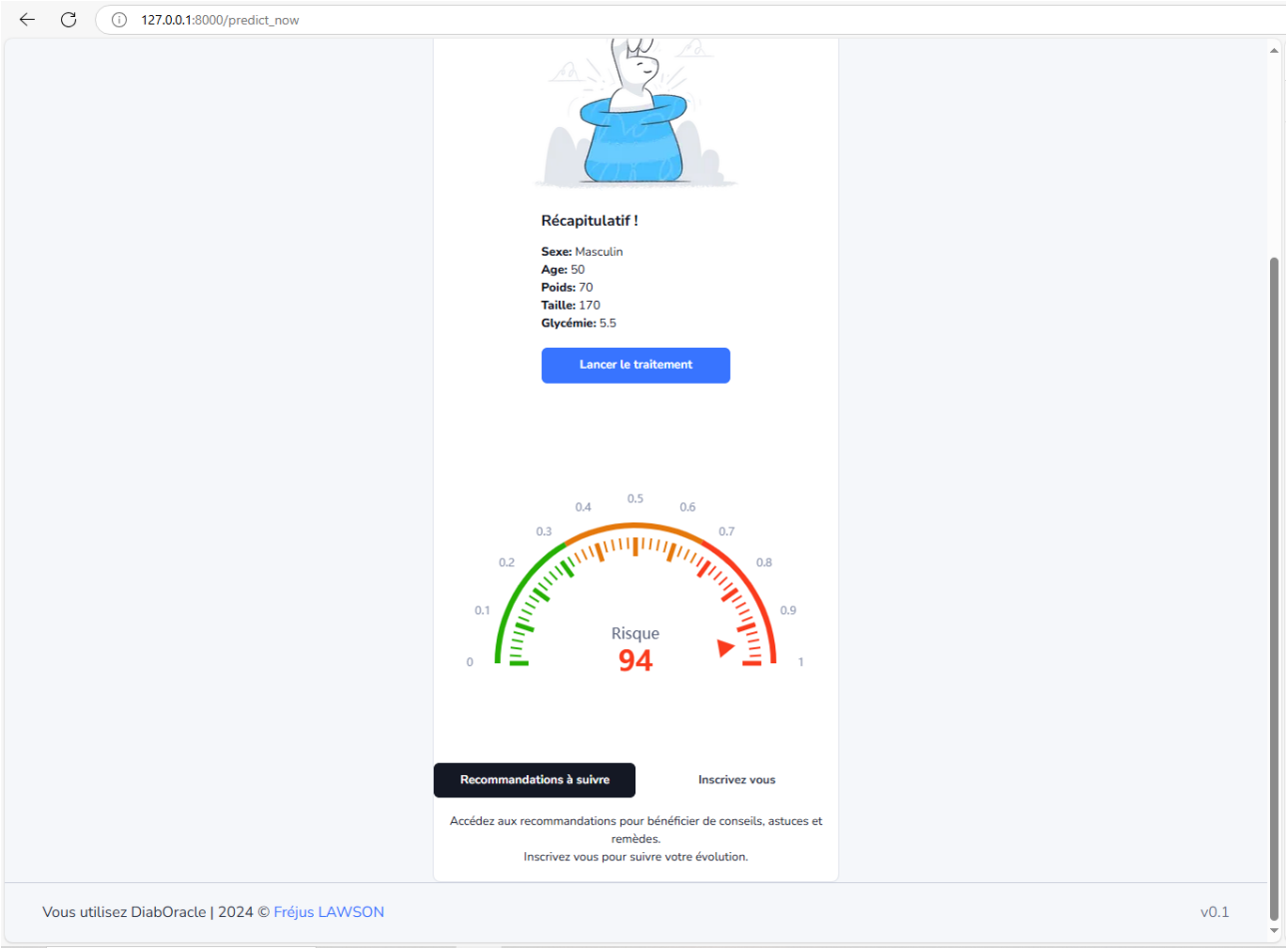


FIGURE 3.3 : Présentation des résultats sous forme de Diagramme

Interprétation des résultats : Pour un individu de 50 ans, pesant 70 kg et mesurant 1,70 m, avec une glycémie de 5,5, il encoure un risque de 94% d'être atteint du diabète.

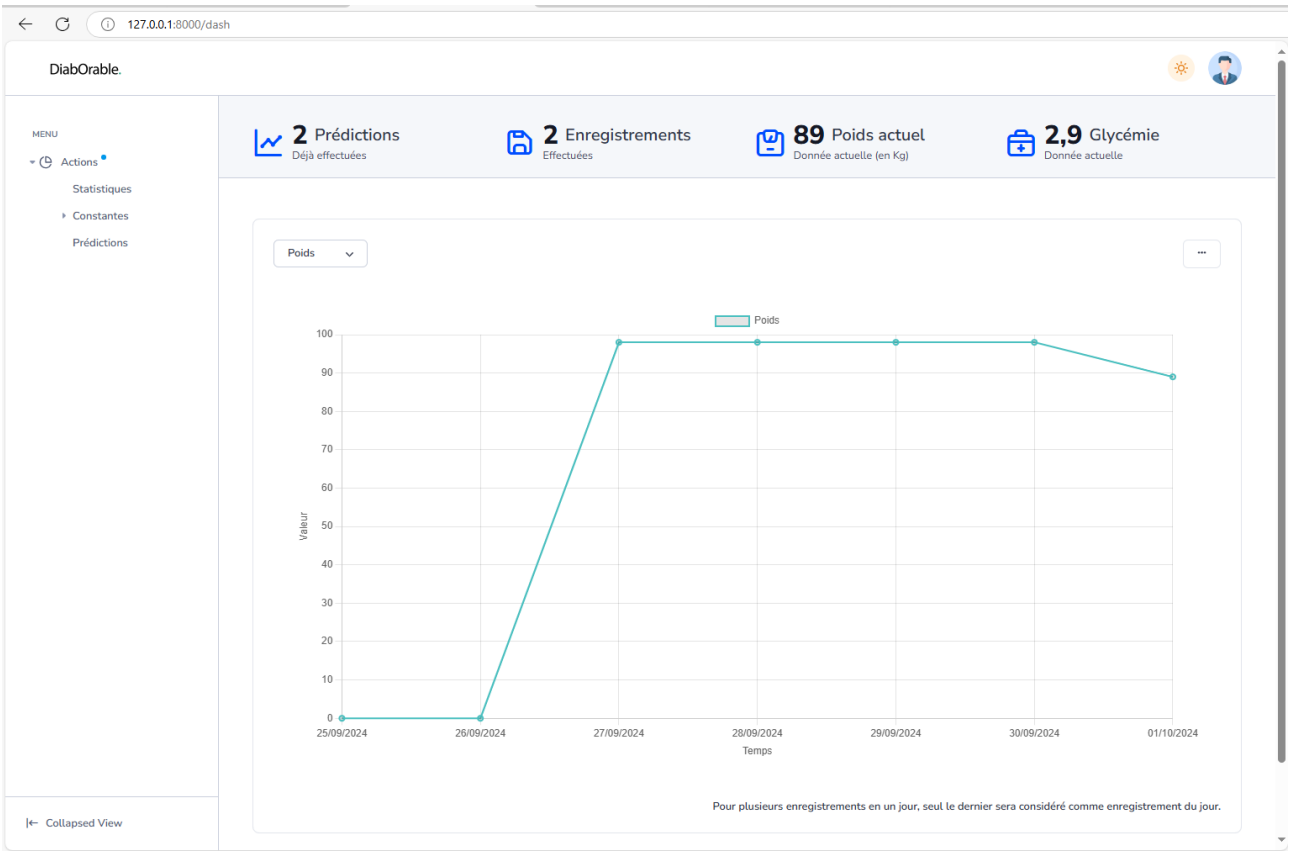


FIGURE 3.4 : Dashbord

Lorsque l'utilisateur se connecte, il a accès à ce tableau de bord où il peut suivre ses constantes dans l'onglet **Statistiques**. Il peut visualiser, à l'aide d'un graphique, l'évolution de son risque, de son poids et de sa glycémie.

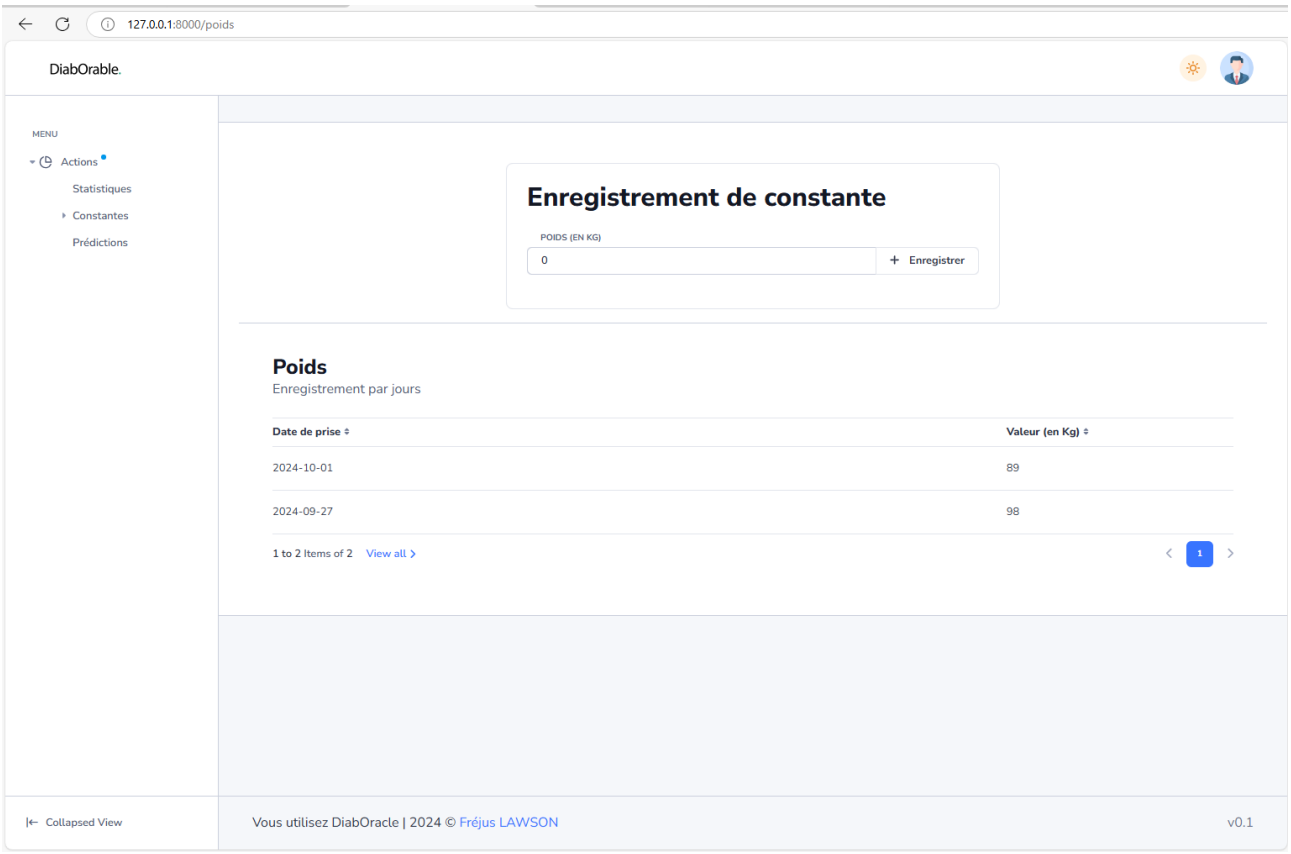


FIGURE 3.5 : Enrégistrement des constantes

Dans l'onglet **Constantes** ci-dessus, les utilisateurs peuvent enregistrer les valeurs de leur poids et de leur glycémie de manière journalière.

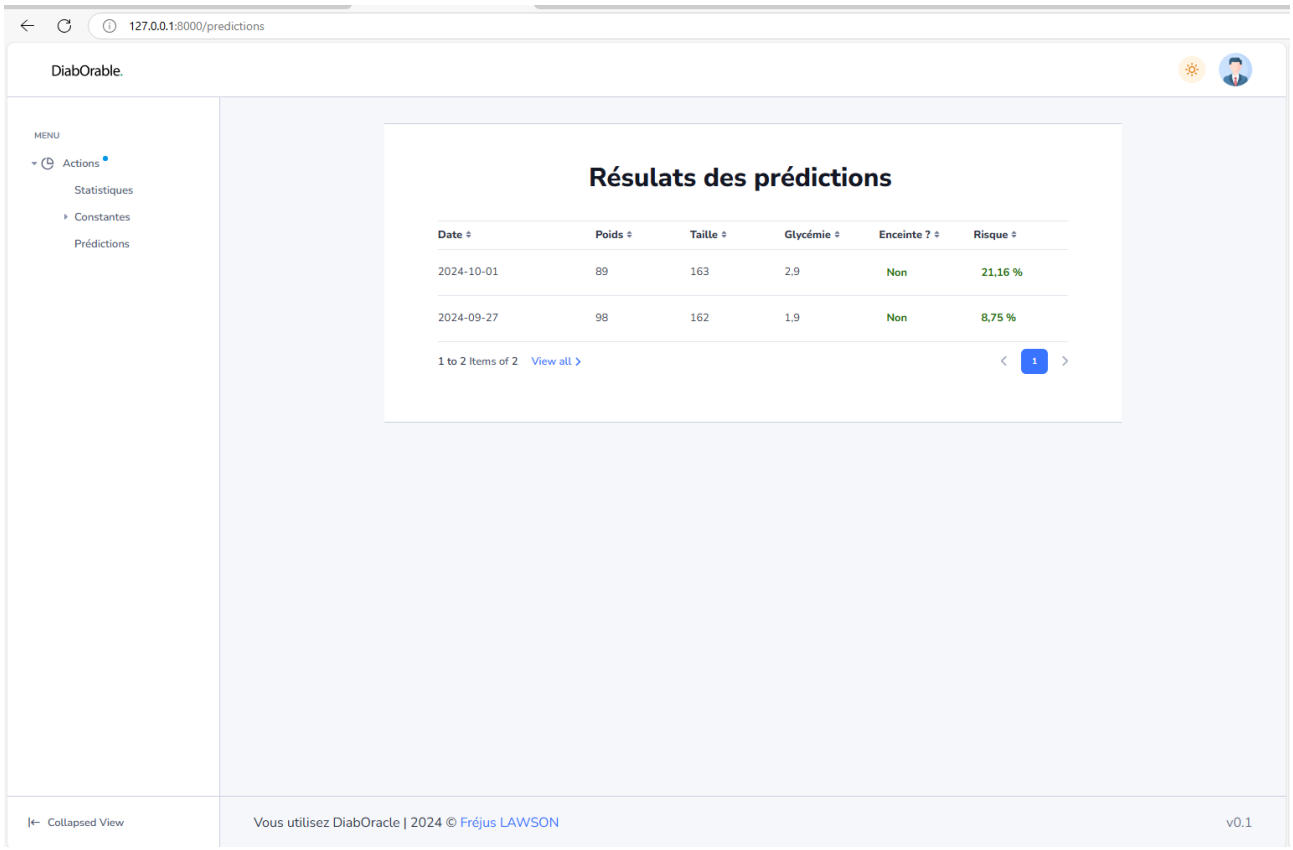


FIGURE 3.6 : Historique des prédictions

L'utilisateur peut consulter les résultats de ses anciennes prédictions (historique) dans l'onglet **Prédictions**.

3.2 Discussion

Nous avons analysé plusieurs applications existantes, telles que *DiabTrend*, *One Drop*, *Diabet AI*, *MD-Calc* et la *Diabetes Australia Risk Calculator*. Chaque application présente des avantages spécifiques, notamment en termes de suivi de la glycémie, de prédiction des risques et de facilité d'utilisation pour les utilisateurs finaux. Cependant, toutes présentent des limites significatives, telles que l'absence de conseils personnalisés et l'inadéquation avec les réalités locales, particulièrement en Afrique. Notre solution, DiabOracle, a été conçue pour répondre à ces défis en intégrant des fonctionnalités innovantes telles que des conseils personnalisés basés sur des données locales, un suivi médical continu avec l'implication des professionnels de santé, et une accessibilité multiplateforme. DiabOracle vise à combler le fossé entre les besoins des utilisateurs africains et les solutions actuellement disponibles sur le marché. En introduisant DiabOracle sur le marché, nous visons à améliorer significativement la prévention et la gestion du diabète, en particulier dans des contextes où les ressources médicales sont limitées et où le diabète constitue un problème de santé publique majeur. L'accessibilité et la pertinence locale de DiabOracle pourraient potentiellement réduire les taux de prévalence du diabète de type 2 en fournissant des outils adaptés et accessibles à tous.

Bien que DiabOracle présente des avantages significatifs dans la gestion et la prédiction du diabète, plusieurs limitations doivent être prises en compte. Tout d'abord, l'acceptation par les professionnels de santé peut être limitée en raison de la nécessité de s'adapter à de nouvelles technologies, ce qui pourrait ralentir l'intégration de DiabOracle dans les processus de soins. De plus, la qualité et la diversité des données d'entraînement sont importantes, car des données non représentatives pourraient entraîner des prédictions inexactes. Il est également important de considérer l'adaptation culturelle et géographique, car les habitudes alimentaires et l'accès aux soins varient entre les régions. La fiabilité des utilisateurs est un autre point critique, car la saisie incorrecte ou irrégulière des données peut affecter la précision des prédictions. Enfin, le modèle prédictif nécessitera des mises à jour régulières pour rester pertinent, impliquant la collecte de nouvelles données et l'intégration des dernières recherches. Malgré ces limitations, DiabOracle offre une base solide pour améliorer la gestion proactive du diabète grâce à des outils numériques, avec des perspectives d'évolution prometteuses pour surmonter ces défis.

3.3 Perspectives d'Amélioration :

- Conception d'un système de chatbot boosté par l'intelligence artificielle pour permettre aux utilisateurs de discuter avec l'IA afin d'en apprendre davantage sur le diabète, et avec leurs médecins pour des conseils personnalisés.
- Développement d'un système de notifications pour améliorer les interactions des utilisateurs avec DiabOracle. Cela inclura des rappels réguliers pour enregistrer les constantes et effectuer des prédictions, ainsi que des informations sur la disponibilité de nouvelles astuces DIY (Do It Yourself).
- Amélioration du modèle prédictif en utilisant des réseaux de neurones pour intégrer les habitudes quotidiennes des utilisateurs, telles que l'activité physique et les habitudes alimentaires.

- Extension des capacités de prédiction de DiabOracle pour inclure les risques d'accidents vasculaires cérébraux (AVC), en analysant des facteurs de risque supplémentaires et en fournissant des conseils préventifs personnalisés.
- Développement d'un modèle d'aide au diagnostic avec propositions de traitement pour les médecins.
- Mise en place d'un modèle économique selon lequel l'accès aux discussions avec l'IA et les professionnels de la santé est régi par un abonnement mensuel pour les patients. Les professionnels de la santé devront également s'abonner pour utiliser le modèle d'aide au diagnostic et pour gérer plusieurs patients à la fois.

Conclusion

En conclusion, cette partie de notre memoire a présenté notre prototype d'application de prédiction du diabète, détaillant ses fonctionnalités principales telles que la saisie des données médicales, la prédiction du risque, et le suivi par des professionnels de la santé, tout en assurant la confidentialité des utilisateurs. Les résultats des tests effectués ont démontré l'efficacité et la fiabilité de ces fonctionnalités, tandis que les perspectives futures soulignent des améliorations potentielles pour enrichir l'expérience utilisateur et étendre les capacités de l'application. Cette présentation confirme le potentiel de l'application à offrir un soutien significatif dans la gestion proactive du diabète, tout en mettant en lumière les opportunités d'évolution pour optimiser ses performances et son impact.

Conclusion Générale

Ce mémoire a exploré le développement d'une application de prédiction du diabète, en se concentrant sur l'état de l'art, la méthodologie de travail et la conception du système, ainsi que la présentation du prototype de l'application. À travers ces grandes étapes, nous avons abordé les enjeux actuels de la gestion du diabète et proposé une solution innovante pour améliorer la prédiction de cette maladie chronique.

La revue de la littérature a permis de situer notre travail dans le contexte des recherches existantes sur la prédiction du diabète. Nous avons identifié les méthodes et technologies employées dans les applications actuelles, ainsi que les lacunes qui persistent dans ce domaine. Cette analyse a révélé l'importance croissante des outils de prédiction basés sur des données et des algorithmes d'apprentissage automatique, tout en mettant en lumière les défis liés à la précision et à la personnalisation des modèles. Les recherches antérieures ont montré que, malgré des avancées significatives, il reste des opportunités pour améliorer les systèmes de prédiction en intégrant des approches plus adaptées aux besoins individuels des patients.

Dans la section méthodologique, nous avons décrit en détail la conception et le développement de notre application. La méthodologie adoptée a inclus la sélection rigoureuse des algorithmes de prédiction ainsi que la préparation des données, comprenant la collecte et le nettoyage des données cliniques pertinentes. La conception du système a été orientée par l'objectif d'intégrer un modèle de prédiction performant tout en assurant une interface utilisateur intuitive et accessible.

Le processus de développement a impliqué des choix technologiques minutieux, notamment le langage de programmation et les frameworks utilisés pour créer une application robuste. La conception a également pris en compte les retours des utilisateurs pour garantir que l'outil soit non seulement fonctionnel, mais aussi convivial. En outre, une attention particulière a été portée à la validation du modèle, avec des techniques telles que la validation croisée pour évaluer sa précision et sa fiabilité.

La présentation du prototype a illustré les résultats concrets de notre travail de recherche et de développement. L'application développée offre une interface claire et conviviale, permettant aux utilisateurs d'entrer leurs données personnelles et de recevoir une évaluation prédictive de leur risque de diabète. Les résultats obtenus montrent que l'application est capable de fournir des prédictions fiables, basées sur des algorithmes sophistiqués et des données actualisées. Les tests effectués ont démontré une précision satisfaisante et une adéquation avec les besoins des utilisateurs finaux.

En conclusion, notre travail a permis de développer une application innovante de prédiction du diabète qui intègre les dernières avancées en matière de traitement des données et d'apprentissage automatique. L'application répond à un besoin crucial de personnalisation et de précision dans la gestion du diabète, et elle a le potentiel d'améliorer significativement la prise en charge préventive

de cette maladie.

Cependant, plusieurs aspects pourraient être améliorés dans les futures versions de l'application. Une expansion des sources de données et une intégration plus poussée avec les systèmes de santé existants pourraient enrichir les capacités prédictives de l'application. De plus, la collecte continue de données et le retour d'expérience des utilisateurs permettront d'affiner les algorithmes et d'optimiser l'interface utilisateur.

En somme, cette recherche ouvre la voie à des développements futurs dans le domaine des applications de santé numérique. La combinaison de technologies avancées et d'une approche centrée sur l'utilisateur promet de transformer la manière dont les maladies chroniques telles que le diabète sont prédite et gérées. Les prochaines étapes incluront l'évaluation de l'impact à long terme de l'application et l'exploration de son intégration dans des contextes cliniques réels.

Bibliographie

[1] Stöcker, Horst and Bosser, Vincent and Marcello, Sandra. "*Toutes les mathématiques et les bases de l'informatique*" 2020

[2] Ludovic Lebart, Alain Morineau, Marie Piron. "*Statistique Exploratoire multidimensionnelle*" 2021

[3] Audibert Thierry, Oussalah Amar. "*Informatique tronc commun*" 2021

[4] Stephane Tufféry. "*Data mining et statistique décisionnelle*" 2012

[5] Jones, J. M. "*Machine Learning for Healthcare*" CRC Press, 2020.

[6] Wang, L., Pedersen, L. G. "*Artificial Intelligence in Healthcare : Prediction of Diseases*" Journal of Healthcare Informatics Research, 2021.

[7] Smith, A. R. "*Improving Patient Outcomes with Machine Learning*" Journal of Medical Systems, 2020.

Webographie

[8] Django Waves, https://djangowaves.com/tips-tricks/fix-import-error-no-module-named-django-core-management/, consulté le 10 Aout 2024.

[9] Apprendre Python, https://apprendrepython.com/comment-installer-des-paquets-python-avec-pip-et-requirements-txt/, consulté le 21 Octobre 2023.

[10] Texmaker, https://www.xm1math.net/texmaker/index_fr.html, consulté le 8 Mars 2024.

[11] Stack Overflow, Communauté et plateforme d'entraide pour développeurs, https://stackoverflow.com, consulté le 8 Septembre 2023.

[12] MedTech Europe, Informations sur les nouvelles technologies médicales, https://medtecheurope.org, consulté le 10 Janvier 2024.

[13] Health IT Analytics, Informations sur les applications d'intelligence artificielle dans la santé, https://healthitanalytics.com, consulté le 20 Février 2024.

[14] Application web, https://digitiz.fr/blog/definition-application-web/,consulté le 20 Février 2024.

[15] Django Software Foundation and individual contributors. Customizing authentication in django. https://docs.djangoproject.com/fr/3.1/topics/auth/customizing/, consulté le 15 juin 2024.

Table des matières

Dédicace ii

Remerciements iii

Résumé iv

. iv

Abstract v

. v

Table des figures vi

Liste des tableaux vii

Liste des acronymes viii

Introduction 1

1 État de l'art 4

Introduction 4

1.1 Présentation des solutions existantes 4

1.2 Comparaison des solutions existantes 10

1.3 Notre Solution 11

Conclusion 11

2 Méthodologie de travail et Conception du système 13

Introduction 13

2.1 Analyse des Besoins 13

2.2 Paramètres de conception 14

2.3 Outils de conception 15

2.4 Conception et Implémentation du modèle prédictif 22

2.4.1 Besoins de données 22

2.4.2 Mesures de performances du modèle 29

2.4.3 Evaluation du risque P 29

2.5 Sécurité 32

2.5.1 Confidentialité des données 32

2.5.2 Intégrité des données 33

2.5.3 Protection contre les menaces externes 33

2.5.4 Conformité légale 33

Conclusion 34

3 Résultats et discussions **35**
Introduction 35
3.1 Présentation du prototype 35
3.2 Discussion 41
3.3 Perspectives d'Amélioration : 41
Conclusion 42

Conclusion **43**

Bibliographie **45**

Webographie **46**

Table des matières **47**