

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE 20 AOUT 55 DE SIKDA
FACULTE DES SCIENCES
DEPARTEMENT d'INFORMATIQUE



Mémoire de fin d'études

En vue de l'obtention du diplôme de Master en Informatique
Spécialité : Réseaux et systèmes distribués

THÈME

Un modèle d'apprentissage pour la prédiction de maladies

Présenté par :

- SEDDIKI Imene
- MEDJEDOUB Meriem

Dirigé par :

Mr. BENOUDINA Lazhar

2021/2022

Dédicace

Nous dédions ce travail à :

Nos parents, frères et sœurs, pour leur appui, leur encouragement, et leur soutien tout au long de notre parcours universitaire, Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible, Merci d'être toujours là pour nous.

Remerciement

Nous tenons à exprimer toute notre gratitude à notre encadreur Mr. BENOUDINA Lazhar pour ses conseils, ses directives et son aide.

Nos remerciements s'adressent aux membres du jury qui ont accepté d'évaluer notre travail. Nous tenons à exprimer nos sincères remerciements à tout le corps professoral et administratif de notre université.

Table de matières

Introduction générale	1
Chapitre 01 : Intelligence Artificielle(AI), et l'AI dans le secteur médicale.	3
1. Introduction	3
2. L'intelligence artificielle	3
2.1 Définition	3
2.2 Historique	4
2.3 Qu'est ce qui contribue al	5
2.4 Application de l'IA.....	5
3. L'intelligence artificielle dans le secteur médicale	6
3.1 Avantages de l'IA dans le secteur médicale	7
3.2 Application de l'IA dans le secteur médicale	7
3.2.1 Développement de médicaments	7
3.2.2 Médecine prédictive.....	8
3.2.3 Médecine préventive.....	9
3.2.4 Robots médicaux	10
3.3 L'IA pour l'aide a décision clinique	10
4. Conclusion.....	11
Chapitre 02 : L'apprentissage automatique (Machine Learning)	12
1. Introduction	12
2. Apprentissage automatique	13
3. Modes d'apprentissage automatique	13
3.1 Apprentissage supervisé	13
3.2.1 Fonctionnement	13
3.2.2 Les problèmes de régression	14
a) Decision Tree (Arbre de décision)	15
a) Random Forest (Forêt d'arbre de décision)	16
a) Comparaison entre les algorithmes de classification DT et RF	17

3.2.3 Les problèmes de classification	18
a) KNN (k-nearest neighbour)	19
a) Naïve Bayes	20
3.2 Apprentissage non supervisé	20
3.3 Apprentissage par renforcement	22
4. Conclusion	22
Chapitre 03 : Conception.	23
1. Introduction	23
2. Architecture et description du système	23
3. Modélisation UML	24
3.1 Description textuelle	24
3.2 Diagramme de cas d'utilisation.....	24
3.2 Diagramme de classe.....	24
3.2 Diagramme de séquence	25
4. Entraînement et test des données	25
4.1 Entraînement des données	26
4.2 Test des données	26
5. Conclusion	27
Chapitre 04 : Implémentation et fonctionnalités.	28
1. Introduction	28
2. Hardware Requis et utilisé :	28
2.1 Hardware requis pour le fonctionnement du système	28
2.2 Hardware utilise pour le développement du système	28
3. Développement	29
3.1 Langages de programmation et environnement de développement	29
3.2 Autres Outils	31
3.2.1 Numpy	31

3.2.2 Matplotlib	32
3.2.3 Pandas.....	32
3.2.4 Seaborn	32
3.2.1 Scikit Learn.....	32
3.2.2 Pickle.....	32
3.3 Dataset	32
4. Interfaces du système	34
5. Conclusion	36
Conclusion générale	37
Bibliographie	38

Liste de figures

Figure [1.1] : Architecture des systèmes experts.

Figure [1.2] : Le rôle de l'IA et L'IoT (Internet Of Things).

Figure [1.3] : l'IA pour l'aide à décision clinique.

Figure [1.3] : Schéma des modèles de l'apprentissage automatique (Machine Learning)

Figure [2.2] : Apprentissage supervisé.

Figure [2.3] : Représentation de problème de régression.

Figure [2.4] : Représentation de l'algorithme d'arbre de décision.

Figure [2.5] : Représentation de Random Forest.

Figure [2.6] : Représentation de problème de classification.

Figure [2.7] : Formule de calcul de distance dans l'algorithme KNN.

Figure [2.8] : Représentation de l'algorithme KNN.

Figure [2.9] : Classification Naïve Bayésienne.

Figure [2.10] : exemple apprentissage non-supervise.

Figure [2.11] : exemple apprentissage non-supervise(2).

Figure [3.1] : Architecture générale du système.

Figure [3.2] : Diagramme de cas d'utilisation.

Figure [3.3] : Diagramme de classes.

Figure [3.4] : Diagramme de séquence.

Figure [4.1]: Ensemble de données d'entraînement (Training data).

Figure [4.2]: Ensemble de données de test (Testing data).

Figure [4.3] : Interface générale du système

Figure [4.4] : Interface de résultat de prédiction (1).

Figure [4.5] : Interface de résultat de prédiction (2).

Figure [4.6] : Interface de prédiction sans saisir le nom.

Liste de tableaux

Table [2.1] : Comparaison entre les algorithmes de classification Decision Tree et Random Forest.

Table [4.1] : Environnement et langages de programmation du système.

Liste des abréviations

IA : Intelligence Artificielle.

DT: Decision Tree.

RF: Random Forest.

Introduction générale

Résumé

L'exploration de données, également appelée découverte de connaissances dans des sources de données, est un domaine de recherche important en informatique. Elle est largement utilisée dans la recherche scientifique (médecine, astronomie, analyse de données biologiques...).

Cette discipline a connu une croissance importante ces dernières années grâce aux avancées récentes en intelligence artificielle et est maintenant appliquée dans plusieurs domaines.

L'objet de réalisation de notre projet est d'exploiter les algorithmes et les méthodes de l'intelligence artificielle pour trouver une façon de prédire les maladies en fonction de symptômes donnés, en utilisant un data set d'un ensemble de données précédentes des patients dans les hôpitaux.

Notre travail démontre l'efficacité de la méthode d'apprentissage automatique dans la médecine prédictive qui ajoute une valeur importante au secteur de santé et aux utilisateurs d'une façon générale.

Problématique

L'avancement parallèle de la médecine et de l'informatique a pu prodiguer des techniques révolutionnaires dans la prédiction, la localisation et même traitement de différentes pathologies. Ainsi, Plusieurs techniques de l'intelligence artificielle inspirées de la biologie telle que les réseaux de neurones et l'apprentissage automatique, sont couramment utilisées dans les applications médicales en vue d'améliorer l'exactitude et la performance des systèmes d'aide au diagnostic médical.

Dans ce contexte, il existe divers problèmes liés aux machines ou appareils qui donneront des résultats erronés ou non précis donc pour obtenir des résultats corrects, nous allons dans ce Project construire un programme qui donnera des prédictions précises basées sur des informations fournies par l'utilisateur (symptômes), et également basées sur les ensembles de données disponibles sur cette machine en exploitant la puissance de l'intelligence artificielle dans ce domaine.

Organisation du projet

Afin d'aboutir à l'accomplissement de notre travail, nous allons suivre le plan suivant :

- ❖ Chapitre 1 : L'Intelligence artificielle, et l'AI dans le secteur médicale.

- ❖ Chapitre 2 : L'apprentissage automatique (Machine Learning).
- ❖ Chapitre 3 : Conception.
- ❖ Chapitre 4 : Implémentation et fonctionnalités.

Chapitre 01 : Intelligence Artificielle(AI), et l'AI dans le secteur médicale

1. Introduction

L'intelligence artificielle comme on la connaît est un domaine qui chaque jour évolue un peu plus en s'appropriant des capacités cognitives humaines, en les développant et parfois même en surpassant ce que les meilleurs êtres humains dans leur domaine sont capables de réaliser.

L'IA S'attache à résoudre des problèmes qui relèvent d'activités humaines de nature variée (perception, prise de décision, planification, diagnostic, interprétation de données, compréhension du langage, conception). Ces problèmes nécessitent de mettre en jeu une grande quantité de données et de connaissances, soit exploitées directement, soit codées sous différentes formes (distributions de probabilités, poids synaptiques, etc.).

Nous présentons dans ce chapitre la définition de l'IA et ses domaines d'application.

2. L'intelligence Artificielle (AI)

2.1 Définition

En 1950, le mathématicien Alan Turing se posait une question : " les machines peuvent-elles penser ? ". En réalité, cette simple interrogation allait bouleverser le monde.

En effet, l'intelligence artificielle a pour but de répondre à la question d'Alan Turing de manière affirmative. Son objectif est de répliquer ou de **simuler l'intelligence humaine dans les machines.**

C'est un but ambitieux, qui soulève aussi de nombreuses questions et suscite le débat. C'est la raison pour laquelle il n'existe **pas encore de définition unique** de l'intelligence artificielle.

La description de " machines intelligentes " ne permet pas d'expliquer ce qu'est véritablement l'intelligence artificielle ou **ce qui rend une machine intelligente**. [14]

L'intelligence artificielle (IA, ou AI en anglais pour *Artificial Intelligence*) consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle. L'IA se retrouve implémentée dans un nombre grandissant de domaines d'application.

2.2 Historique

La notion voit le jour dans les années 1950 grâce au mathématicien **Alan Turing**. Dans son livre *Computing Machinery and Intelligence*, ce dernier soulève la question d'apporter aux machines une forme d'intelligence. Il décrit alors un test aujourd'hui connu sous le nom « Test de Turing » dans lequel un sujet interagit à l'aveugle avec un autre humain, puis avec une machine programmée pour formuler des réponses sensées. Si le sujet n'est pas capable de faire la différence, alors la machine a réussi le test et, selon l'auteur, peut véritablement être considérée comme « intelligente ». Le développement de l'IA a connu de nombreuses fluctuations.

Dès le milieu des **années 60**, la recherche autour de l'IA sur le sol américain était principalement financée par le Département de la Défense. Dans le même temps, des laboratoires ouvrent çà et là à travers le monde. Certains experts prédisaient à l'époque que « *des machines seront capables, d'ici 20 ans, de faire le travail que toute personne peut faire* ». Si l'idée était visionnaire, même en 2018 l'intelligence artificielle n'a pas encore pris cette importance dans nos vies.

En **1974** arrive une période appelée le « *AI Winter* ». Beaucoup d'experts ne réussissent pas à faire aboutir leurs projets et les gouvernements britannique et américain réduisent leurs financements pour les académies. Ils préfèrent soutenir des idées ayant plus de chances de déboucher sur quelque chose de concret.

Dans les **années 80**, le succès des systèmes experts permet de relancer les projets de recherche sur l'intelligence artificielle. Un système expert était un ordinateur capable de se comporter comme un expert (humain), mais dans un domaine bien précis. Grâce à ce succès, le marché de l'IA atteint une valeur d'un milliard de dollars, ce qui motive les différents gouvernements à de nouveau soutenir financièrement plus de projets académiques.

Le développement exponentiel des performances informatiques, notamment en suivant la loi de Moore, permet entre **1990 et 2000** d'exploiter l'IA sur des terrains jusqu'alors peu communs. On retrouve à cette époque le data mining, ou encore les diagnostics médicaux. Il faudra attendre **1997** pour une véritable sortie médiatique

lorsque le fameux Deep Blue créé par IBM a battu Garry Kasparov, alors champion du monde d'échec. [8]

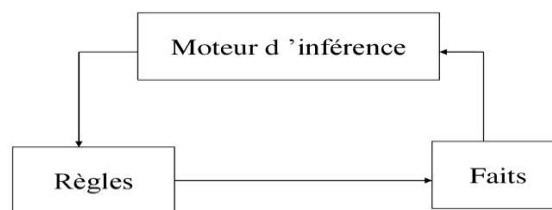
2.3 Qu'est-ce qui contribue à l'IA ?

L'IA est une technologie basée sur des disciplines telles que l'informatique, la linguistique, la psychologie, la biologie, l'ingénierie et les mathématiques. L'IA est principalement basée sur le développement des technologies informatiques associées à l'intelligence humaine, telles que l'apprentissage, le raisonnement et la résolution de problèmes. Parmi les objectifs de l'IA on a :

- **La création des systèmes experts** : Un système expert est un programme informatique qui simule l'action et le comportement d'un être humain ou d'une organisation possédant des connaissances et des expériences étendues dans un domaine particulier.

Intelligence Artificielle

Architecture d'un système expert



07/04/2019

Intelligence Artificielle

14

Figure [1.1] : Architecture des systèmes experts [9]

- Faits : état initial, état désiré.
- Règle : Base de connaissance.
- Moteur d'inférence

- **L'implémentation de l'intelligence humaine dans les machines** : L'IA est un moyen efficace permettant aux ordinateurs et aux logiciels à penser en utilisant des systèmes experts pour adopter un comportement intelligent et conseiller les utilisateurs en appliquant de l'apprentissage et de la réflexion.

2.4 Application de l'IA

L'intelligence artificielle a été dominante dans divers domaines tels que :

- L'intelligence artificielle de jeu : joue un rôle crucial dans les jeux stratégiques où la machine peut penser à un grand nombre de positions possibles basées sur une connaissance heuristique, tels que le poker, les échecs etc.
- Les systèmes experts : certaines applications intègrent des machines et des logiciels permettant de raisonner et de proposer des conseils aux utilisateurs.
- Les robots intelligents : les robots sont des machines ou des appareils fonctionnant automatiquement ou par télécommande. Ils ont des capteurs pour détecter les données physiques du monde réel telles que la température, la chaleur, la lumière, les mouvements, le son, les chocs, la pression etc. Ils sont capables d'apprendre de leurs erreurs et de s'adapter à un nouvel environnement afin d'accomplir les tâches confiées à un humain.
- La reconnaissance vocale : certains systèmes intelligents sont capables d'entendre et de comprendre la voix humaine captée. Ces systèmes permettent d'analyser le langage en termes de phrases et de leurs significations pendant qu'un humain le parle.
- La reconnaissance de l'écriture manuscrite est l'un des domaines de recherche les plus fascinants et les plus difficiles dans le domaine du traitement de l'image et de la reconnaissance des formes au cours des dernières années. Les applications dotées d'un système de reconnaissance de l'écriture manuscrite contribuent énormément à l'avancement du processus d'automatisation et peuvent améliorer l'interface entre l'homme et la machine.
- Les systèmes de vision : ces systèmes interprètent et comprennent les entrées visuelles sur l'ordinateur. Par exemple.
- **Les secteurs médicaux utilisent les systèmes experts pour laide à la décision clinique.**
- Un avion espion prend des photographies qui sont utilisées pour comprendre des informations spatiales.
- La police utilise un logiciel informatique capable d'arrêter une personne criminel grâce aux systèmes de la reconnaissance faciale.

3. L'intelligence artificielle dans le secteur médicale

L'intelligence artificielle est partout, mais elle trouve plus particulièrement des applications intéressantes dans le domaine de la santé. Les données médicales constituent une ressource inestimable pour prédire des maladies, diagnostiquer une pathologie ou améliorer le suivi des patients.

L'intelligence artificielle ne va pas remplacer les professionnels de la médecine ; elle va être le stéthoscope du 21^e siècle et une collaboration fructueuse entre l'homme et la technologie pourrait nous apporter le changement positif dans la médecine que nous souhaitons tous si fortement.

La santé numérique nous donnera plus de données sur la santé que jamais auparavant, et l'IA nous aidera à les analyser pour trouver de nouvelles façons de traiter les maladies, de réduire les tâches administratives, de rationaliser les pratiques médicales, d'optimiser les horaires des médecins et des patients...etc.

3.1 Avantages de L'IA dans le secteur médicale

Des expériences centrées sur l'utilisateur

Se basant sur de vastes ensembles de données et sur l'apprentissage automatique, les organisations de santé peuvent trouver des informations plus rapidement et plus précisément grâce à l'IA. Elles peuvent ainsi mieux satisfaire les intervenants en interne et les patients qu'elles servent.

Amélioration de l'efficacité des opérations

En examinant les modèles de données, les technologies d'IA aident les organisations du secteur de la santé à tirer le meilleur parti de leurs données, de leurs actifs et de leurs ressources, en augmentant l'efficacité et en améliorant les performances des flux de travaux cliniques et opérationnels, des processus et des opérations financières.

Connexion de données de santé disparates

Les données de santé sont souvent fragmentées et se présentent sous différents formats. Avec l'IA et les technologies d'apprentissage automatique, les organisations peuvent connecter des données disparates pour obtenir une image plus unifiée des individus qui se cachent derrière les données.

3.2 Application de l'IA dans le domaine médicale

3.2.1 Développement de médicaments

L'intelligence artificielle se déploie progressivement dans le monde de la sante et principalement au niveau de l'optimisation de développement de nouveaux traitements.

- **Optimisation du Drug design**

Aujourd'hui, la plupart des nouveaux médicaments sont souvent développés via ce que l'on appelle le Drug design faisant appel à la chimie combinatoire. L'intelligence artificielle apporte de nouvelles perspectives en permettant de déterminer quelles sont les substances les plus intéressantes à tester sur le plan pharmacologique.

Dans le cadre de la recherche de candidat médicament, on observe plusieurs étapes de développement notamment au niveau du choix des principes actifs. Le plus souvent on utilise la chimie combinatoire. Face au volume de données à traiter, l'intelligence artificielle apparaît comme une solution pour optimiser la recherche.

- **L'IA créatrice de médicaments**

L'intelligence artificielle optimise les phases amont de recherche mais permet également le développement de nouvelles molécules dans des délais réduits.

Une première molécule mise au point grâce à l'intelligence artificielle va être testée dans les prochaines semaines sur des humains au Japon pour permettre de traiter les Troubles obsessionnels compulsifs (TOC). Développée par Sumitomo Dainippon Pharma en collaboration avec Exscientia, cette IA a permis d'accélérer le processus de développement pour le réduire de 5 ans à 12 mois.[1]

3.2.2 Médecine prédictive

L'essor de la panoplie d'outils qui permettent de générer et d'analyser les données de santé d'un grand nombre d'individus offre de nouvelles perspectives cliniques. Les développements de la génomique ont d'ores et déjà permis, notamment grâce au séquençage du génome, d'identifier des mutations ou variantes génétiques qui favorisent ou qui sont à l'origine de tumeurs, et par conséquent d'optimiser les stratégies thérapeutiques associées [2]. La connaissance de ces particularités génétiques et moléculaires tout comme celles d'autres facteurs de risques liés aux habitudes de vie et à l'environnement permettent soit d'améliorer le suivi et le dépistage du cancer chez les patients à risque, soit de proposer des traitements personnalisés [3], [4]. L'analyse de toutes ces données qui contiennent de plus en plus d'informations longitudinales, permet, grâce à l'utilisation d'algorithmes de modélisation, d'affiner les paramètres menant au diagnostic, de définir le bénéfice des traitements et

d'estimer, voire prédire, les durées et les conditions de survie de patients à risque [5].

3.2.3 Médecine préventive

L'IA, avec l'utilisation d'appareils médicaux connectés (IoT), joue un grand rôle pour garantir que les gens restent en bonne santé.

- **Pour les patients** : L'IA aide déjà les patients à être plus autonome dans leur parcours de soins et à améliorer leur qualité de vie au quotidien. Un nombre croissant de personnes choisissent d'adopter des objets connectés pour suivre leurs statistiques de santé au quotidien. La collecte et l'analyse de ces données de santé peuvent offrir une perspective unique sur la santé des individus et de la population en général, permettant ainsi de se diriger vers une médecine plus préventive.
- **Pour les médecins** : La technologie de l'IA donne aux professionnels de la santé un aperçu des habitudes et des besoins quotidiens de leurs patients. Cela leur permet de les guider, les informer et les soutenir au mieux. La vaste quantité de données générées et collectées ne peut être traitée par l'humain uniquement. L'utilisation d'algorithmes d'IA permet d'économiser beaucoup de temps et d'efforts humains, et donc de rendre les processus plus efficaces. Les soignants peuvent se concentrer uniquement sur des tâches à forte valeur ajoutée.
- **Pour les chercheurs** : La collecte de données de santé provenant de sources multiples permet de faire avancer la recherche médicale. Les données provenant de diverses sources sont mises en commun et agrégées afin d'accéder à des statistiques de santé en temps réel.

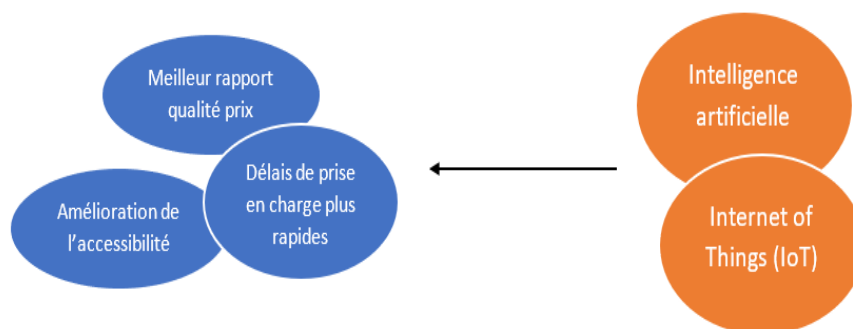


Figure [1.2] : Le rôle de l'IA et L'IoT (Internet Of Things)

3.2.4 Robots médicaux

- **Chirurgie assistée par robot** : Les robots chirurgicaux utilisent l'IA pour utiliser les informations des opérations précédentes afin d'améliorer les techniques chirurgicales. Les données des dossiers préopératoires sont intégrées aux mesures opérationnelles pour améliorer les résultats de l'opération. Ces opérations sont peu invasives et la précision de l'instrument assisté par robot permet de réduire le degré du traumatisme post-op.
- **Chirurgies robotisées autonomes** : Bien qu'elles soient actuellement limitées à la science-fiction, les opérations robotisées peuvent devenir une réalité à l'avenir. L'utilisation du machine Learning pour combiner la reconnaissance des formes motrices et l'interprétation visuelle des données peut permettre d'étendre la dextérité du chirurgien aux robots et de faire de la chirurgie robotique autonome une réalité. La chirurgie robotisée se limite actuellement à la commande à distance des robots par le chirurgien via un ordinateur, mais cela pourrait se transformer dans le futur.
- **Robots auxiliaires** : Ce sont des robots qui trouvent une application dans divers domaines, notamment les soins aux patients, les soins infirmiers, et les soins aux personnes âgées et aux patients affaiblis.
- **Chatbots** : Les Chatbots sont des algorithmes alimentés par l'IA capables d'effectuer des conversations avec les patients. Ils ont le potentiel de devenir le premier point de contact pour les soins de santé primaires. La gravité de la requête est déterminée et les Chatbots peuvent soit résoudre le problème, soit le transmettre au médecin. L'utilisation généralisée des Chatbots réduit considérablement la charge de travail du médecin et évite d'avoir à se rendre inutilement chez les professionnels de santé.

3.3 IA pour l'aide à la décision clinique

Inondés de volumes gigantesques de données sur la santé et croulant sur des responsabilités de plus en plus nombreuses, les cliniciens peinent à trouver le temps de se tenir au courant des dernières études médicales tout en continuant à assurer des soins centrés sur le patient.

En appliquant les technologies d'apprentissage automatique aux dernières données biomédicales et aux dossiers médicaux électroniques, les professionnels de la santé

peuvent rapidement extraire des informations précises, pertinentes et factuelles préalablement organisées par des homologues. Certains outils d'aide à la décision clinique basés sur l'IA proposent un traitement en langage naturel et ont reçu un entraînement à la discipline concernée.

Les utilisateurs peuvent ainsi rédiger leurs questions comme ils le feraient s'ils interrogeaient un collègue dans une conversation quotidienne, et recevoir des réponses rapides et fiables. [6]

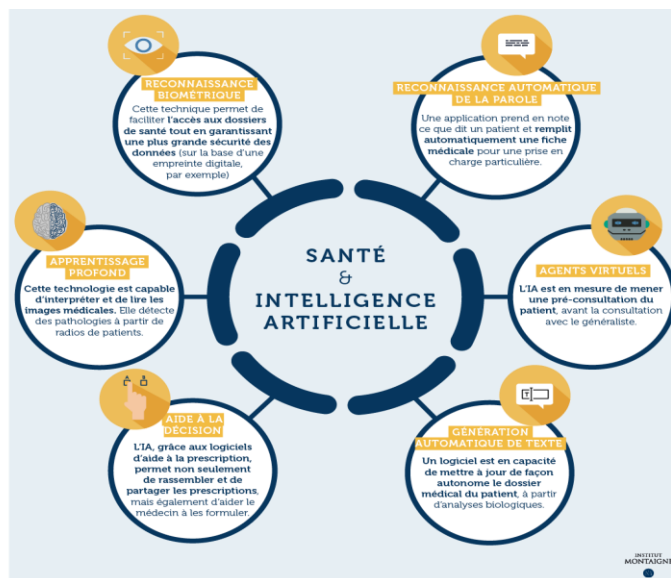


Figure [1.3] : l'IA pour l'aide à décision clinique

4. Conclusion

Dans ce chapitre on a présente comment l'introduction de l'IA et de l'apprentissage automatique dans le secteur de la santé a apporté de nombreux avantages aux organisations de santé et aux patients.

Chapitre 02 : Apprentissage automatique (Machine Learning)

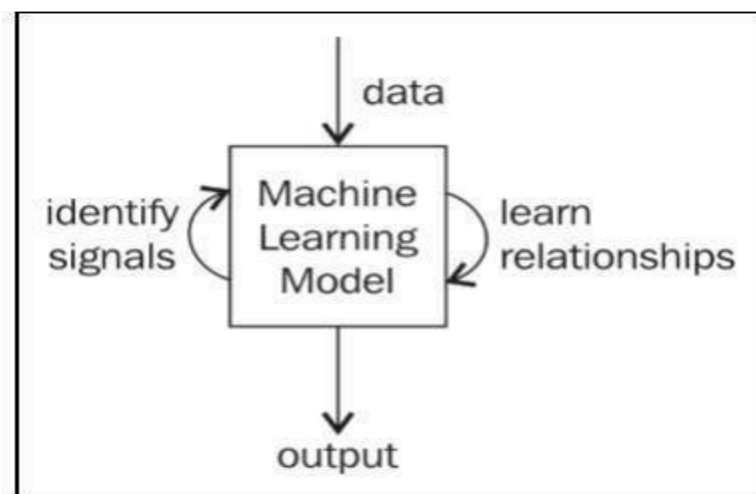
1. introduction

De manière générale, un programme informatique tente de résoudre un problème pour lequel nous avons la solution. Par exemple : calculer la moyenne générale des étudiants, classer les étudiants selon leur moyenne. . .

Pour certains problèmes, nous ne connaissons pas de solution exacte et donc nous ne pouvons pas écrire de programme informatique. Par exemple : reconnaître automatiquement des chiffres écrits à la main à partir d'une image scannée, déterminer automatiquement une typologie des clients d'une banque, jouer automatiquement aux échecs contre un humain ou un autre programme. . .

En revanche, pour ces problèmes il est facile d'avoir une base de données regroupant de nombreuses instances du problème considéré. L'apprentissage automatique consiste alors à programmer des algorithmes permettant d'apprendre automatiquement de données et d'expériences passées, un algorithme cherchant à résoudre au mieux un problème considéré.

Alors dans ce chapitre on va présenter l'approche d'apprentissage automatique (Machine Learning).



An overview of machine learning models

Figure [2.1] : Schéma des modèles de l'apprentissage automatique (Machine Learning)

2. Apprentissage automatique

L'apprentissage automatique (AA) "**Machine Learning**" est à la croisée de plusieurs disciplines :

- **Les statistiques** : pour l'inférence de modèles à partir de données.
- **Les probabilités** : pour modéliser l'aspect aléatoire inhérent aux données et au problème d'apprentissage.
- **L'intelligence artificielle** : pour étudier les tâches simples de reconnaissance de formes que font les humains (comme la reconnaissance de chiffres par exemple), et parce qu'elle fonde une branche de l'AA dite symbolique qui repose sur la logique et la représentation des connaissances.
- **L'optimisation** : pour optimiser un critère de performance afin, soit d'estimer des paramètres d'un modèle, soit de déterminer la meilleure décision à prendre étant donné une instance d'un problème.
- **L'informatique** : puisqu'il s'agit de programmer des algorithmes et qu'en AA ceux-ci peuvent être de grande complexité et gourmands en termes de ressources de calcul et de mémoire.

3. Modes d'apprentissage

3.1 Apprentissage Supervisé

L'apprentissage supervisé est le paradigme d'apprentissage **le plus populaire** en Machine Learning et en Deep Learning. Comme son nom l'indique, cela consiste à superviser l'apprentissage de la machine en lui montrant des **exemples** (des données) de la tâche qu'elle doit réaliser. Les applications sont nombreuses : Reconnaissance vocale, vision par ordinateur, régressions, classifications... La grande majorité des problèmes de Machine Learning et de Deep Learning utilisent **l'apprentissage supervisé**.

3.1.1 Fonctionnement :

Avec l'apprentissage supervisé, la machine peut apprendre à faire une certaine tâche en étudiant des **exemples** de cette tâche.

Exemple : elle peut apprendre à reconnaître une photo de chien après qu'on lui ait montré des millions de photos de chiens. Ou bien, elle peut apprendre à traduire le français en chinois après avoir vu des millions d'exemples de traduction français-chinois

D'une manière générale, la machine peut apprendre une **relation** $f : x \longrightarrow y$ qui relie x à y en ayant **analysé** des millions d'exemples d'associations.

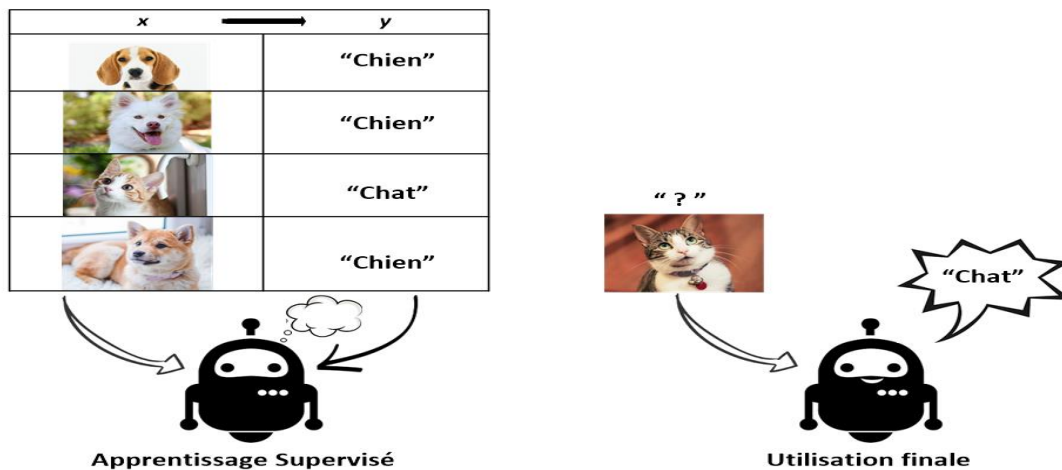


Figure [2.2] : Apprentissage supervisé. [7]

L'apprentissage supervisé fonctionne en 4 étapes :

1. Importer un **Dataset (x, y)** qui contient nos exemples.
2. Développer un **Modèle** aux paramètres aléatoires.
3. Développer une **Fonction Coût** qui mesure les erreurs entre le modèle et le Dataset.
4. Développer un **Algorithme d'apprentissage** pour trouver les **paramètres** du modèle qui **minimisent la Fonction Coût**.

Les **applications** de l'Apprentissage Supervisé sont nombreuses, très nombreuses ! On peut les diviser en 2 catégories de problèmes : les **Régressions**, et les **Classifications**.

3.1.2 Les problèmes de Régression :

Les problèmes de **Régression** correspondent aux situations dans lesquelles la machine doit prédire la valeur d'une **variable quantitative** (variable continue)

Exemples de variables y :

- le prix d'un appartement
- l'évolution du climat
- le cours de la bourse
- durée de vie d'un patient

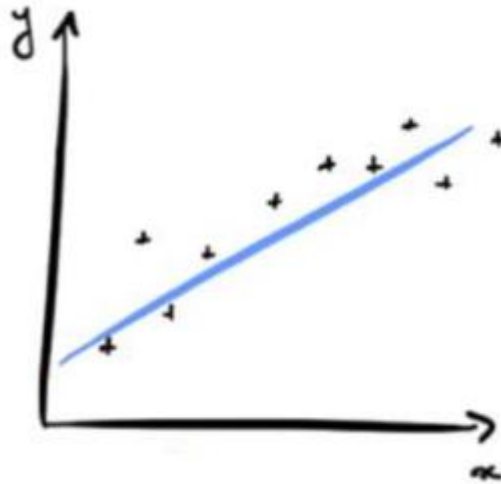


Figure [2.3] : Représentation de problème de régression.

Dans l'apprentissage automatique, les algorithmes de régression tentent d'estimer la fonction (f) des variables d'entrée (x) aux variables de sortie numériques ou continues (y). Par exemple, lorsque vous disposez d'un ensemble de données sur les maisons et que vous êtes invité à prédire leurs prix, il s'agit d'une tâche de régression car le prix sera une sortie (valeur) continue. Les exemples d'algorithmes de régression courants incluent **Linear Regression**, **Decision Tree**, **Random Forest** et **Support Vector Regression**.

a) **Decision Tree (Arbre de décision):**

L'apprentissage de l'arbre de décision aussi appelé *Decision Tree* (DT) en anglais, est une technique d'apprentissage automatique supervisé pour induire un arbre de décision à partir des données d'apprentissage. Un arbre de décision est un arbre binaire (arbre dans lequel chaque nœud a deux nœuds enfants). Dans les structures arborescentes, les feuilles représentent des classes (également appelées étiquettes de classe)

Il peut être utilisé soit pour la classification ou bien pour la régression) Pour la classification, chaque feuille d'arbre est marquée par une étiquette de classe, plusieurs feuilles peuvent avoir la même étiquette.

Pour la régression, une valeur est également affectée à chaque feuille de l'arbre, de sorte que la fonction d'approximation est constante. Pour atteindre un nœud feuille afin d'obtenir une réponse du vecteur d'entité en entrée, la procédure de prédiction commence par le nœud racine.

À partir de chaque nœud non feuille, la procédure va vers la gauche (sélectionne le nœud enfant gauche comme nœud observé suivant) ou vers la droite en fonction de la valeur d'une certaine variable dont l'index est stocké dans le nœud observé.

L'arbre de décision est un algorithme itératif qui, à chaque itération, va séparer les individus en k groupes (généralement $k=2$ et on parle d'arbre binaire) pour expliquer la variable cible.

La première division (on parle aussi de *split*) est obtenue en choisissant la variable explicative qui permet la meilleure séparation des individus. Cette division donne des sous-populations correspondant au premier nœud de l'arbre.

Le processus de split est ensuite répété plusieurs fois pour chaque sous-population (nœuds précédemment calculés) jusqu'à ce que le processus de séparation s'arrête.

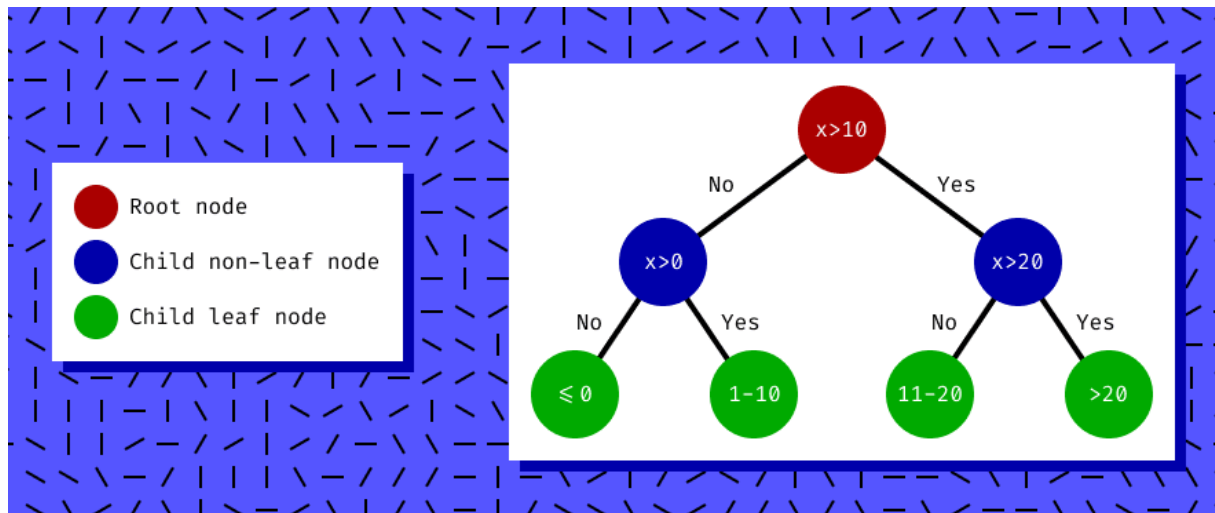


Figure [2.4] : Représentation de l'algorithme de classification L'arbre de décision.

- **Root node:** Noeud Racine.
- **Child non-leaf node:** nœud non feuille.
- **Child leaf node:** nœud enfant feuille.

b) Random Forest (Forêt d'arbre de décision):

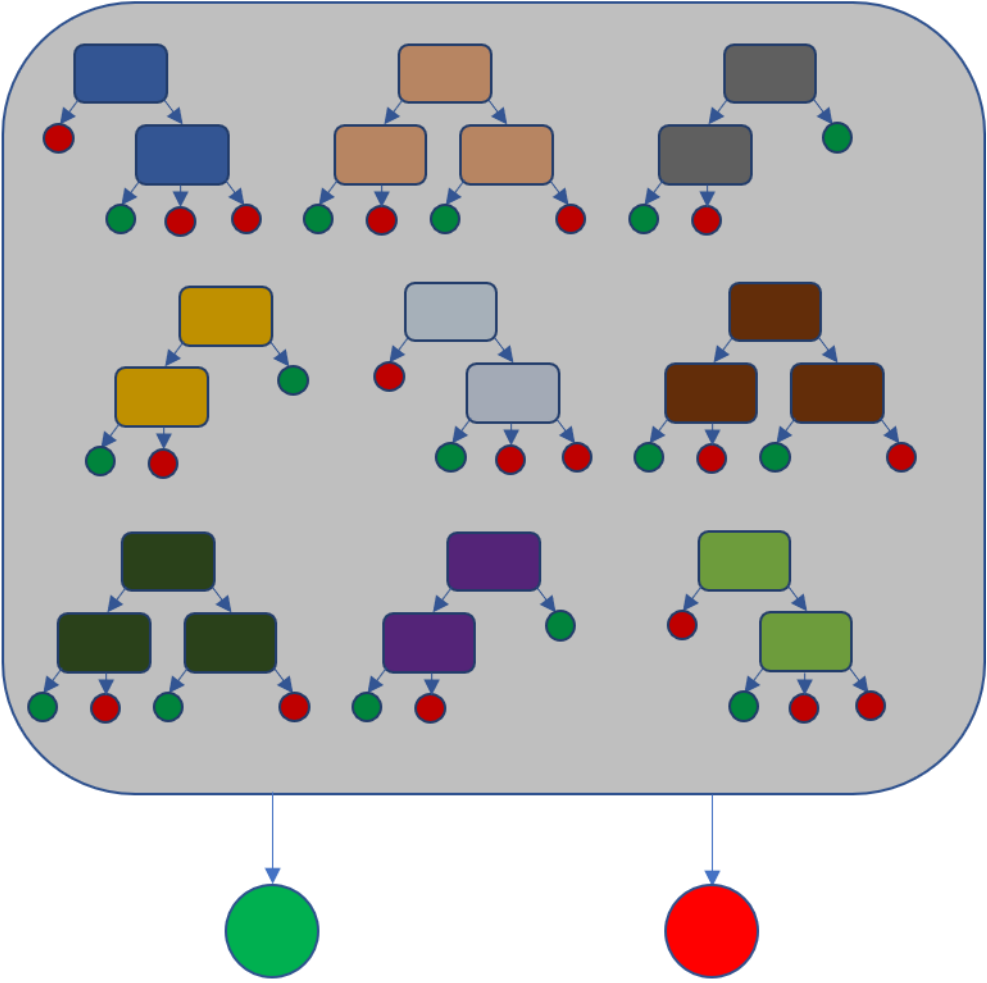
Les forêts aléatoires ou bien *Random Forest* (RF) en anglais, sont parmi les algorithmes qui sont utilisés dans la classification supervisée.

Cet algorithme est proposé par Breiman (2001) au début des années 2000.

Il s'agit de l'un des algorithmes qui restent efficaces lorsqu'il est appliqué à de grandes quantités de données.

Les forêts aléatoires combinent des prédicateurs ou des estimateurs des **arbres**, donnant naissance à ce que l'on appelle maintenant les arbres de décision. Plus généralement, ce sont des algorithmes établis dont leur principe général est de construire une collection de prédicateurs afin d'agréger ensuite toutes leurs prédictions. En classification, l'agrégation consiste par exemple à faire un vote majoritaire parmi les étiquettes de classe fournis par les prédicateurs. Concernant la

régression, la cible de régression prédite d'un échantillon d'entrée est calculée comme la cible de régression prédite moyenne des arbres dans la forêt.



Random Forest

Figure [2.5] : Représentation de Random Forest

c) Comparaison entre les algorithmes de classification DT et RF:

Decision Tree	Random Forest
Un outil de décision qui utilise un graphe arborescent, ou un model de décisions et leurs conséquences possible, y compris les résultats déventements fortuits, les couts des ressources et l'utilité.	Une méthode d'apprentissage d'ensemble qui fonctionne en construisant une multitude d'arbres de décision à des moments de formation et en mettant la classe en fonction des arbres individuels.
Il ya une possibilité de sur-ajustement.	Réduit le risque de sur-ajustement.
Donne des résultats moins précis.	Donne des résultats plus précis.
Plus simple et plus facile à comprendre, interpréter et visualiser.	Plus complexe.

Table [2.1] : Comparaison entre les algorithmes de classification Decision Tree et Random Forest

3.1.3 Les problèmes de Classification

Les problèmes de **Classification** correspondent aux situations dans lesquelles la machine doit prédire la valeur d'une **variable qualitative** (variable discrète). Autrement dit, la machine doit **classer** ce qu'on lui donne dans des classes.

Exemples de variables y :

- Email Spam / non Spam
- Cancer / non Cancer
- Photo de Chat / Chien

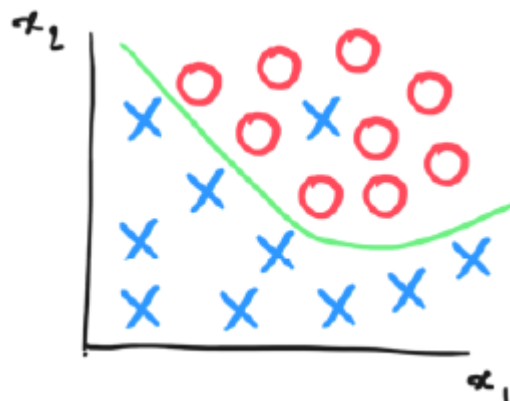


Figure [2.6] : Représentation de problème de classification. [7]

On peut ainsi regrouper des données dans des clusters (c'est le **Clustering**), détecter des **anomalies**, ou encore **réduire la dimension** de données très riches en compilant les dimensions ensemble.

Les algorithmes de classification tels que SVM (Support Vector Machine), K-NN, **Naïve bayes** sont des algorithmes pour classer des données. Ces données peuvent être des mots, des couleurs, des sons, etc. Dans un contexte de classification d'images, ces algorithmes peuvent être utilisés comme classificateurs basés sur des informations textuelles.

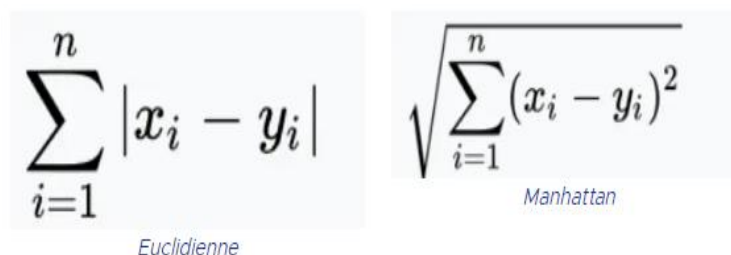
a) KNN (k-nearest neighbors) :

L'algorithme des K plus proches voisins ou K-nearest neighbors (KNN) est un algorithme de Machine Learning qui appartient à la classe des algorithmes d'apprentissage supervisé simple et facile à mettre en œuvre qui peut être utilisé pour résoudre les problèmes de classification et de régression.[15]

L'intuition derrière **l'algorithme des K** plus proches voisins est l'une des plus simples de tous les algorithmes de Machine Learning supervisé :

Étape 1 : Sélectionnez le nombre K de voisins

Étape 2 : Calculez la distance


$$\sum_{i=1}^n |x_i - y_i|$$

Euclidienne

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Manhattan

Figure [2.7] : formule de distance dans l'algorithme KNN. [15]

Du point non classifié aux autres points.

Étape 3 : Prenez les K voisins les plus proches selon la distance calculée.

Étape 4 : Parmi ces K voisins, comptez le nombre de points appartenant à chaque catégorie.

Étape 5 : Attribuez le nouveau point à la catégorie la plus présente parmi ces K voisins.

Étape 6 : Notre modèle est prêt :

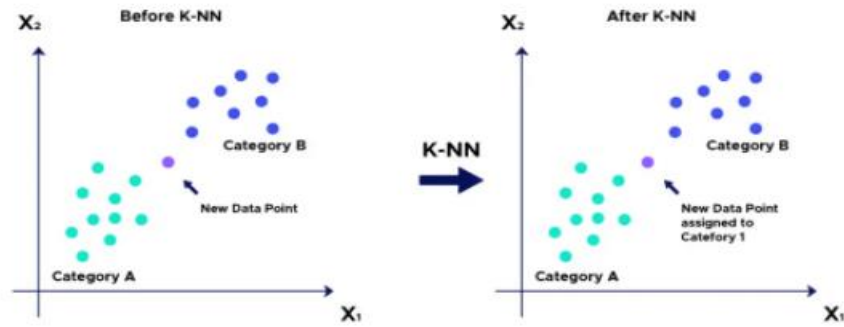


Figure [2.8] : Représentation de l'algorithme KNN. [15]

b) Naïve Bayes:

L'algorithme *Naïve Bayes* (NB) est un algorithme d'apprentissage supervisé basé sur l'application du théorème de Bayes avec l'hypothèse d'indépendance conditionnelle entre chaque paire de caractéristiques pour calculer les probabilités d'appartenance à une classe. Le théorème de Bayes est défini par la relation suivante :

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

L'algorithme *Naïve Bayes* suppose que y est une variable de classe et X est un vecteur d'entités dépendant (de taille n) où: $X = x_1, x_2, \dots, x_n$. Bien que, l'algorithme Naïve Bayes est simple dans sa conception et son développement. L'utilisation de cet algorithme a donné des résultats très encourageants.

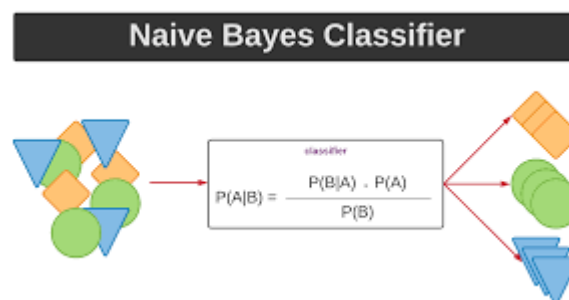


Figure [2.9] : Classification Naïve Bayésienne.

3.2 Apprentissage non supervisé :

Une autre méthode d'apprentissage pour développer des programmes de Machine Learning est l'**apprentissage non Supervisé** (Unsupervised Learning). Cette méthode est utilisée quand notre Dataset ne contient pas d'exemples qui indiquent ce que l'on cherche.

On dispose d'un ensemble d'objets sans aucune valeur cible associée ; il faut apprendre un modèle capable d'extraire les régularités présentes au sein des objets pour mieux visualiser ou appréhender la structure de l'ensemble des données

Exemple :

ON regroupe en 2 familles selon leur ressemblance

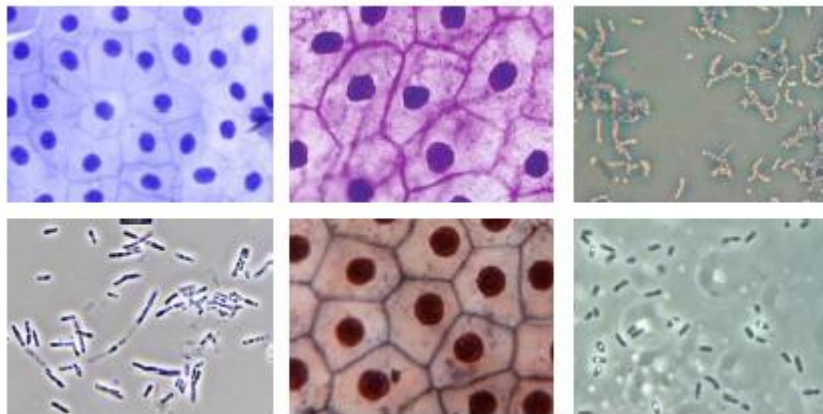


Figure [2.10] : Exemple Apprentissage non-supervisé.

Dans l'apprentissage non-supervisé, on dispose ainsi d'un Dataset x sans variable y et la machine **apprend** à reconnaître des **structures** dans les données x qu'on lui montre.

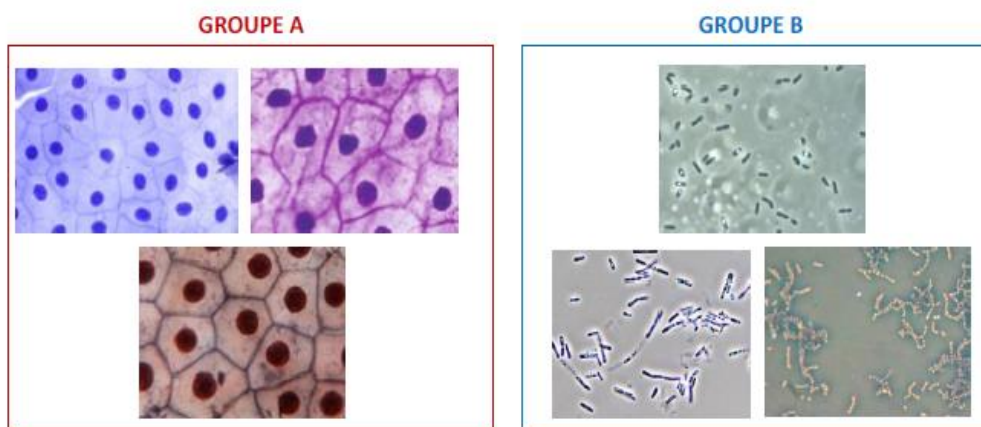


Figure [2.11] : Exemple Apprentissage non-supervisé. [7]

3.3 Apprentissage par renforcement

L'apprentissage par renforcement consiste à laisser la machine apprendre à faire une tâche (par exemple piloter un mini drone) en la laissant pratiquer seule. Quand la machine réussit ce qu'elle entreprend, elle reçoit un **bonus**. Quand elle échoue, elle reçoit un **malus**.

On dispose d'un ensemble de séquences de décisions (politiques ou stratégiques) dans un environnement dynamique, et pour chaque action de chaque séquence une valeur de récompense (la valeur de récompense de la séquence est alors la somme des valeurs des récompenses des actions qu'elle met en œuvre) ; il faut apprendre un modèle capable de prédire la meilleure décision à prendre étant donné un état de l'environnement.[7]

4. Conclusion

Dans ce chapitre on a présenté l'approche d'apprentissage automatique, sa définition et les modes d'apprentissage automatique en se basant sur les algorithmes utilisés pour la réalisation de ce projet.

Chapitre 03 : Conception du système prédiction de maladies

1. Introduction

Notre travail consiste à implémenter un système de détection de maladies basé sur l'apprentissage automatique, et comme tout projet informatique, ce projet nécessite une phase d'implémentation, donc dans ce chapitre on va présenter les différents diagrammes qui modélisent notre système ainsi que l'architecture générale et la description du système.

2. Architecture et description du système

La prédiction de maladies à l'aide de l'apprentissage automatique prédit la présence d'une maladie pour l'utilisateur basé sur les symptômes donnés.

L'architecture de notre système se compose de divers ensembles de données à travers lesquels nous comparerons les symptômes de l'utilisateur, puis ses ensembles de données sont transformés en plus petits ensembles, et à partir de là et à l'aide des algorithmes de classification tels que Decision Tree (Arbre de décision) et Random Forest on classe ces derniers ensembles, ensuite traités par les technologies d'apprentissage automatique et intégrés au modèle de prédiction de maladies.

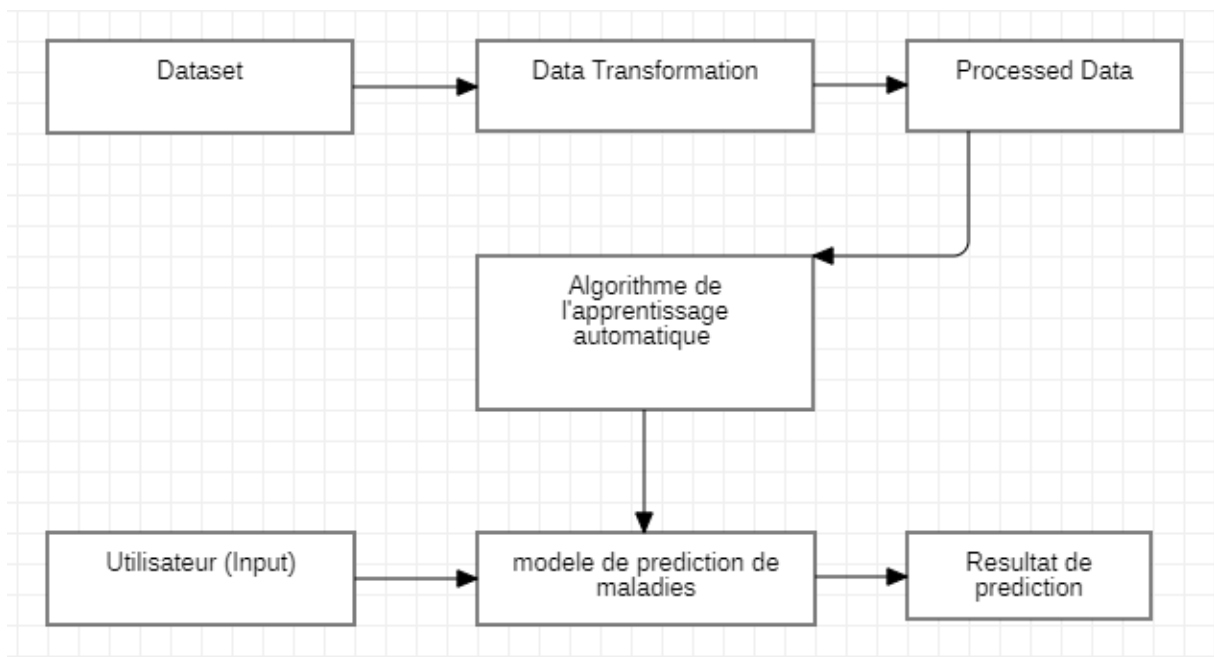


Figure [3.1] : Architecture générale du système

3. Modélisation UML

3.1 Description textuelle

Système : Prédiction de maladies.

Scénario : Une fois l'utilisateur est sur les systèmes, il introduit son nom d'abord puis il commence à choisir les symptômes (Plus de symptômes plus de précision moins de symptômes moins de précision), Pour prédire sa maladie il doit cliquer sur le bouton "PREDICT" pour actualiser la page il clique sur le bouton "RESET", le bouton RESET annule tout action précédentes.

3.2 Diagramme de cas d'utilisation

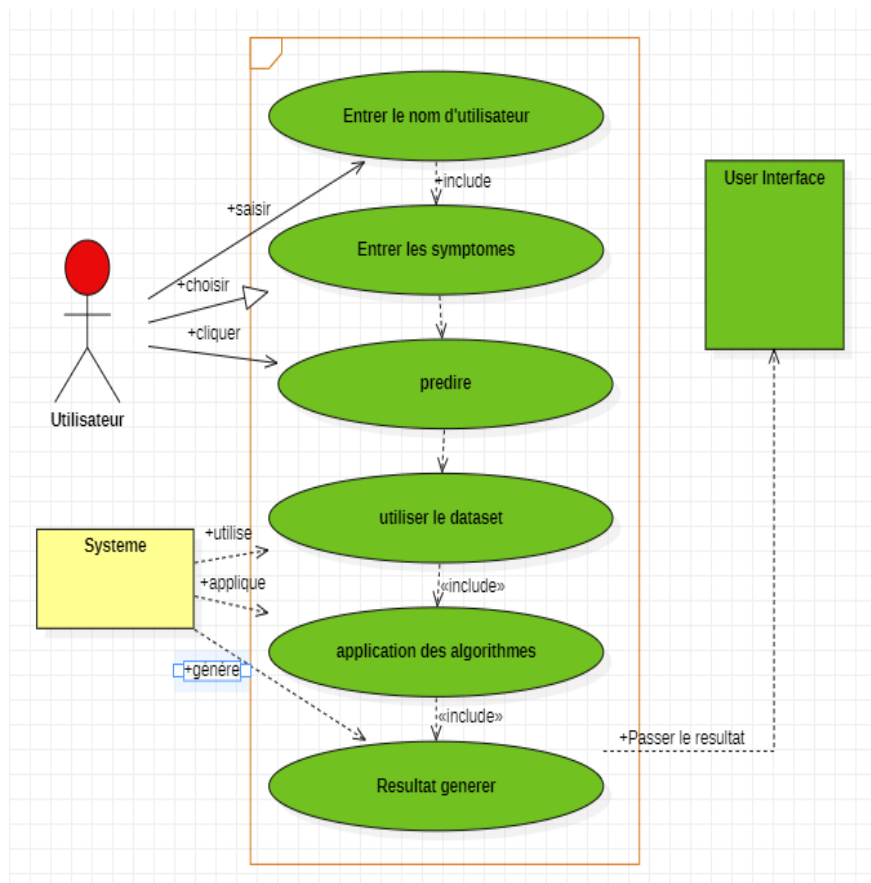


Figure [3.2] : Diagramme de cas d'utilisation

3.3 Diagramme de classes

Le diagramme de classes contient les informations concernant les classes du système, ainsi que les attributs et méthodes associées à chaque classe.

Toutes ces informations sont nécessaires pour la conception de notre système.

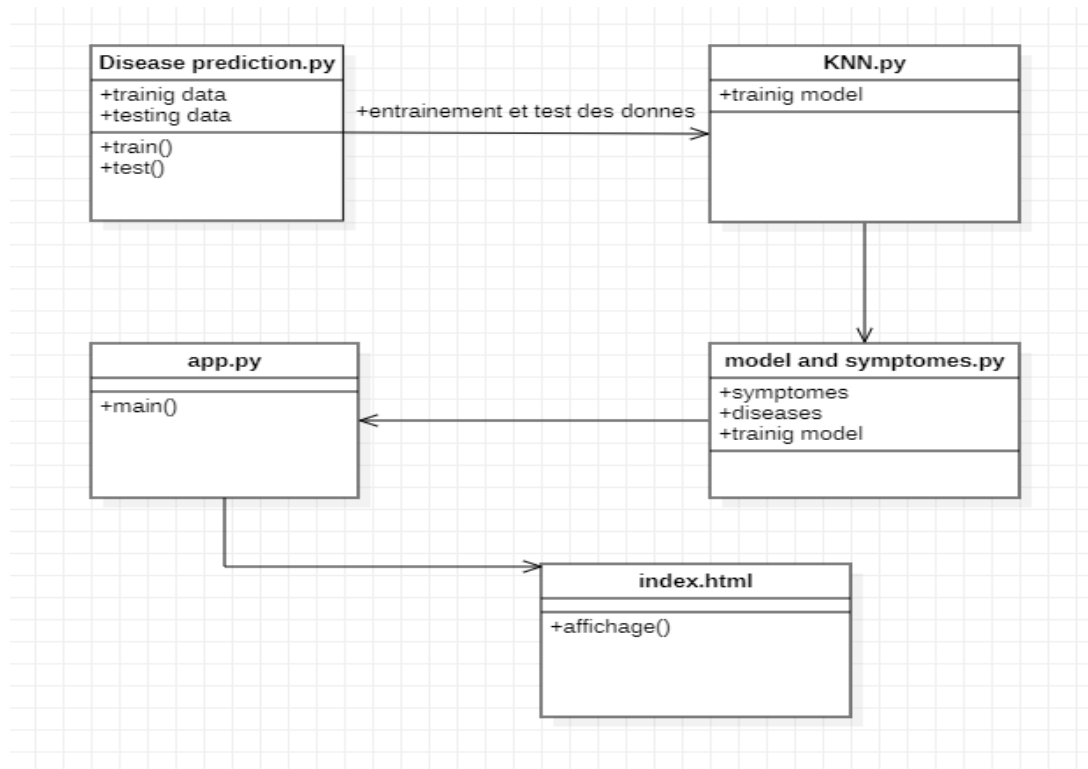


Figure [3.3] : Diagramme de classe.

3.4 Diagramme de séquence

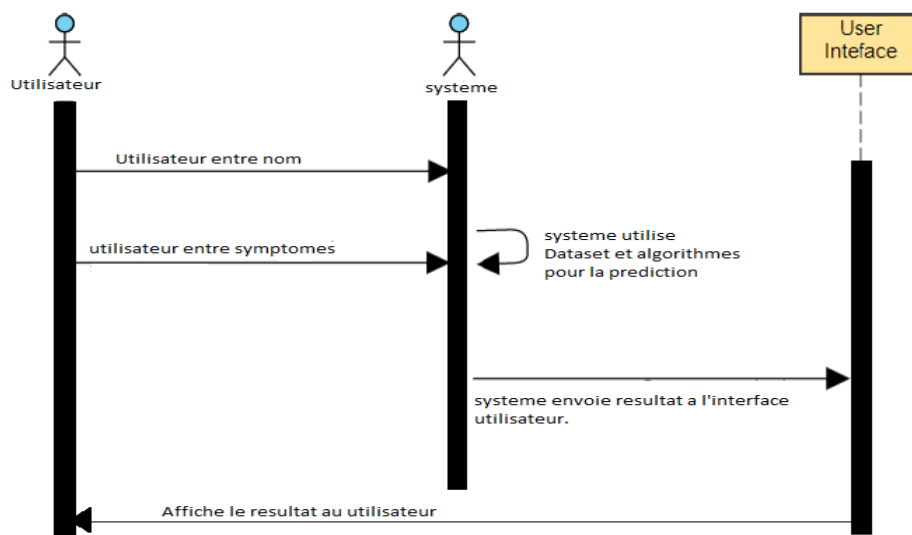


Figure [3.4] : Diagramme de séquence.

4. Entrainement et test de données

4.1 Entraînement des données

L'un des principes pris en compte par l'apprentissage automatique est d'utiliser les données pour prendre des décisions. Concrètement, les techniques d'apprentissage automatique construisent des modèles ou des fonctions à l'aide de données présentées. Ce processus est appelé l'entraînement (Training) des données, et l'ensemble de données utilisé est l'ensemble d'apprentissage. La formation consiste à sélectionner un modèle ou une fonction qui « correspond » le mieux à l'ensemble de formation. En ce qui concerne la classification, l'ensemble d'apprentissage est une collection d'exemples étiquetés.

C'est-à-dire que l'ensemble d'apprentissage est un sous-ensemble de $X \times Y$

Cette situation est connue sous le nom d'apprentissage supervisé, puisque les données d'apprentissage **supervisent** le processus d'apprentissage.

Soit F une famille de classificateurs, indexés par un paramètre \mathbf{z} qui prend valeurs dans un jeu de paramètres \mathbf{Z} . Par exemple, F peut être la famille des classificateurs linéaires, et $\mathbf{Z} = (\mathbf{w}, \mathbf{b}) \in X \times \mathbf{R}$.

Soit l'ensemble d'entraînement donné par :

$$D = \{(x^{(i)}, y^{(i)}) \in X \times Y : 1 \leq i \leq n\}$$

Pour $n \in \mathbf{N}$. Soit $fit : \mathbf{Z} \rightarrow \mathbf{R}$ mesurer l'ajustement aux données ; c'est-à-dire que $fit(\mathbf{z})$ donne l'ajustement entre l'ensemble d'apprentissage D et le classificateur défini par le paramètre \mathbf{Z} . Ensuite la tâche d'entraînement se résume au problème d'optimisation

$$\text{Optimiser }_{z \in \mathbf{Z}} fit(\mathbf{z}).$$

Que l'optimisation soit une maximisation ou une minimisation dépend de l'interprétation de la fonction d'ajustement ($fit()$).

Trouver une bonne mesure d'adéquation entre un classificateur et un ensemble d'apprentissage est d'une importance capitale.

4.2 Test de données

Dans l'entraînement de données, nous faisons de notre mieux pour construire un classificateur qui correspond aux données d'entraînement étiquetée. Cependant, notre objectif ultime est d'avoir un classificateur qui fonctionne bien sur toutes les données appropriées, pas seulement sur les exemples particuliers de l'ensemble d'apprentissage.

Nous utilisons le terme généralisation pour désigner les performances d'un classificateur sur de nouvelles données inédites. Étant donné que l'efficacité d'un classificateur réside dans sa capacité à faire des prédictions utiles même sur de nouvelles données, il est important d'étudier sa généralisation.

Nous mesurons la généralisation en rassemblant une nouvelle collection de données étiquetées, que nous appelons l'ensemble de test (Testing Data), et reporter éventuellement plusieurs quantités numériques qui présentent à quel point le classificateur a pu prédire les étiquettes dans l'ensemble de test.

a) Précision (Accuracy) :

La précision est le pourcentage de prédiction correcte dans le test de données, le pourcentage le plus haut indique la meilleure performance.

Dans ce projet, la majorité des classificateurs donnent un résultat de 100 % de précision, alors on a choisi KNN classificateur comme modèle d'entraînement.

5. Conclusion

Dans ce chapitre, nous avons tenté de mettre en lumière la manière de concevoir et de réaliser un système de prédiction de maladie

Afin d'assurer une implémentation correcte et réussie, nous avons utilisé UML comme langage de modélisation.

On a commencé par une architecture générale du système, ensuite les diagrammes (UML) de cas d'utilisation, de classe et de séquences.

Chapitre 04 : Implémentation et fonctionnalités.

1. Introduction

Le projet Prédiction des maladies à l'aide de l'apprentissage automatique est développé pour faire repérer les maladies générales à des stades précoces.

Le projet "Prédiction des maladies à l'aide de l'apprentissage automatique" est implémenté entièrement en python.

L'utilisateur a besoin du nom et doit sélectionner les symptômes dans le menu déroulant donné, pour un résultat plus précis, l'utilisateur doit entrer tous les symptômes donnés, puis le système fournira le résultat précis.

Cette prédiction est essentiellement effectuée à l'aide de Random Forest, Decision Tree et Naïve Bayes.

Nous avons utilisé ces 3 algorithmes pour fournir une image plus claire des résultats et l'utilisateur doit être satisfait de son résultat prédit.

Dans ce chapitre on va présenter les langages de programmation, environnement de développement ainsi que tous les outils utilisés qui sont nécessaire pour le développement de ce système.

Ensuite, nous détaillons les principales interfaces graphiques et fonctionnalités de notre système

2. Hardware Requis et utilisé

2.1 Hardware Requis pour le fonctionnement du système

- ✓ Processeur : Pentium 4, Intel Core i3, i5, i7 and 2 GHz Minimum
- ✓ RAM: 1GB ou plus.
- ✓ Hard Disk : 20 GB ou plus
- ✓ Matériel Input : Clavier et souris.
- ✓ Matériel Output : Ecran / Pc.

2.2 Hardware utilisé pour le développement du système

- ✓ Processeur: Intel Core(TM) i7 2.10GHz. (Hp Elite Book Windows 10)
- ✓ RAM: 4GB.

3. Développement :

3.1 Langage de programmation et Environnement de développement :

- **Front End:** HTML/CSS
- **Back End:** Python Flask
- **Environnement :** Jupyter notebook



Html/CSS :

Le HTML est un langage qui a été créé en 1991. Les sigles « HTML » sont l'abréviation de « HyperText Markup Language » ou « langage de balisage hypertexte » en français.

Le HTML est donc un langage de balisage, c'est-à-dire un langage qui va nous permettre de définir les différents contenus d'une page.

Le CSS a été créé en 1996, soit 5 ans après le HTML. Les sigles « CSS » sont l'abréviation de « Cascading StyleSheets » ou « feuilles de styles en cascade » en français.

Le CSS vient résoudre un problème bien différent du HTML : en effet, le HTML sert à définir les différents éléments d'une page, à leur donner du sens. Le CSS, lui, va servir à mettre en forme les différents contenus définis par le HTML en leur appliquant des styles. Le HTML va donc créer la structure des pages tandis que le CSS va nous permettre de modifier l'apparence des contenus de la page. On va ainsi par exemple pouvoir définir la taille, la couleur ou l'alignement de certains contenus HTML et notamment en l'occurrence de certains textes dans notre page. [10]



Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels.

En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font.

Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages. [11]



Flask est un micro-Framework python facile et simple qui permet de faire des applications web évolutives. Flask dépend de la boîte à outils WSGI de Werkzeug et du moteur de templates Jinja.

Un micro framework est un framework qui tente de fournir uniquement les composants absolument nécessaires à un développeur pour créer une application. Dans le cas des frameworks d'applications Web, un micro framework peut être spécifiquement conçu pour la construction d'API pour un autre service ou une autre application.

Le micro dans le micro framework signifie que Flask vise à garder le code de base simple mais extensible. Flask ne prendra pas beaucoup de décisions pour vous, par exemple quelle base de données utiliser. Les décisions qu'il prend, telles que le moteur de templates à utiliser, sont faciles à modifier. Tout le reste est à vous, de sorte que Flask puisse répondre à tous vos besoins et à tous ce que vous ne voulez pas en même temps.

En définissant seulement le moteur de templates et un système de routes, Flask vous laisse le choix de personnaliser (en ajoutant des packages ou en développant les vôtres) pour la gestion des formulaires par exemple, vous avez donc la main sur


	votre code. [12]
	<p>Le nom Jupyter Rassemble les noms de ces trois langages de programmation : Julia, Python et R.</p> <p>Jupyter Notebook est une application client-serveur.</p> <p>Elle permet la création et le partage de documents Web au format JSON constitués d'une liste ordonnée de cellules d'entrées et de sorties et organisés en fonction des versions successives du document.</p> <p>Les cellules peuvent contenir, entre autres, du code, du texte au format Markdown, des formules mathématiques ou des contenus médias.</p> <p>Le traitement se fait avec une application client fonctionnant par Internet, à laquelle on accède par les navigateurs habituels. Il est nécessaire pour cela que soit installé et activé dans le système le serveur Jupyter Notebook. Les documents Jupyter créés peuvent s'exporter aux formats HTML, PDF, Markdown ou Python par exemple, ou bien se partager par email, avec Dropbox, Git Hub ou un lecteur Jupyter Notebook. [13]</p>

Table [4.1] : Environnement et langages de programmation.

3.2 Autres outils

3.2.1 Numpy:

Numpy est un paquet de traitement de tableaux à usage général. Il fournit des objets de tableaux multidimensionnels de haute performance et des outils pour travailler avec ces tableaux. Numpy est un conteneur efficace de données multidimensionnelles génériques. Il facilite les opérations mathématiques sur les tableaux et leur vectorisation. Cela améliore considérablement les performances et accélère le temps d'exécution.

3.2.2 Matplotlib:

Matplotlib est la bibliothèque de traçage pour Python qui fournit une API orientée objet pour intégrer des tracés. Grâce à Matplotlib, on peut tracer des histogrammes, graphes de fonctions...Etc.

3.2.3 Pandas:

Pandas est une bibliothèque écrite pour le langage de programmation Python pour la manipulation et l'analyse de données. En particulier, elle propose des structures de données et des opérations de manipulation de tableaux numériques et de séries chronologiques.

3.2.4 Seaborn:

Seaborn est une bibliothèque Python de visualisation de données basée sur Matplotlib. Elle fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants et informatifs.

3.2.5 Scikit-learn:

Scikit-learn (Sklearn) est la bibliothèque la plus utile et la plus robuste pour l'apprentissage automatique en Python. Il fournit une sélection d'outils efficaces pour l'apprentissage automatique et la modélisation statistique, notamment la classification, la régression, le regroupement et la réduction de la dimensionnalité via une interface de cohérence en Python.

3.2.6 Pickle:

Pickle en Python est principalement utilisé pour la sérialisation et la dé-sérialisation d'une structure d'objet Python. En d'autres termes, c'est le processus de conversion d'un objet Python en un flux d'octets pour le stocker dans un fichier/base de données, maintenir l'état du programme à travers les sessions ou transporter des données sur le réseau.

3.3 Dataset :

L'ensemble de data est collecté de **Kaggle.com**

L'ensemble de données complet se compose de deux fichiers CSV, un pour l'entraînement de données (Training) et l'autre pour le test (Testing).

Chaque fichier CSV comporte 133 colonnes. 132 de ces colonnes sont des symptômes qu'une personne éprouve et la dernière colonne est le pronostic.

Ces symptômes sont mappés à 42 maladies dans lesquelles vous pouvez classer ces ensembles de symptômes.

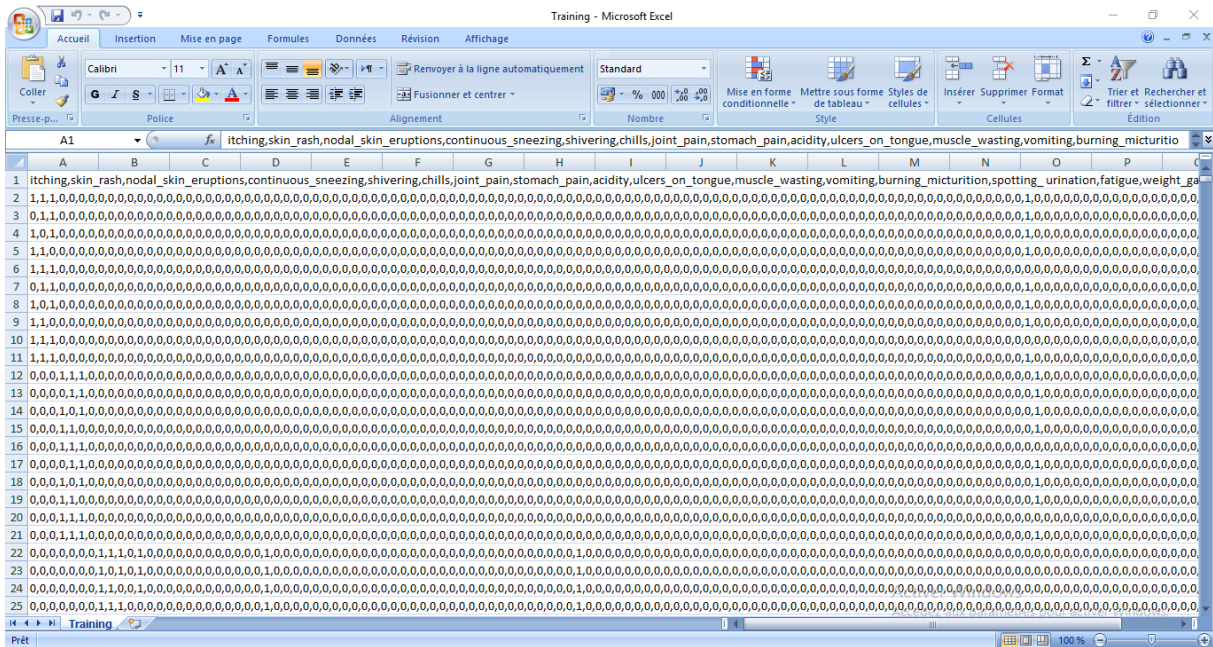


Figure [4.1] : Ensemble de données d'entrainement Training Data

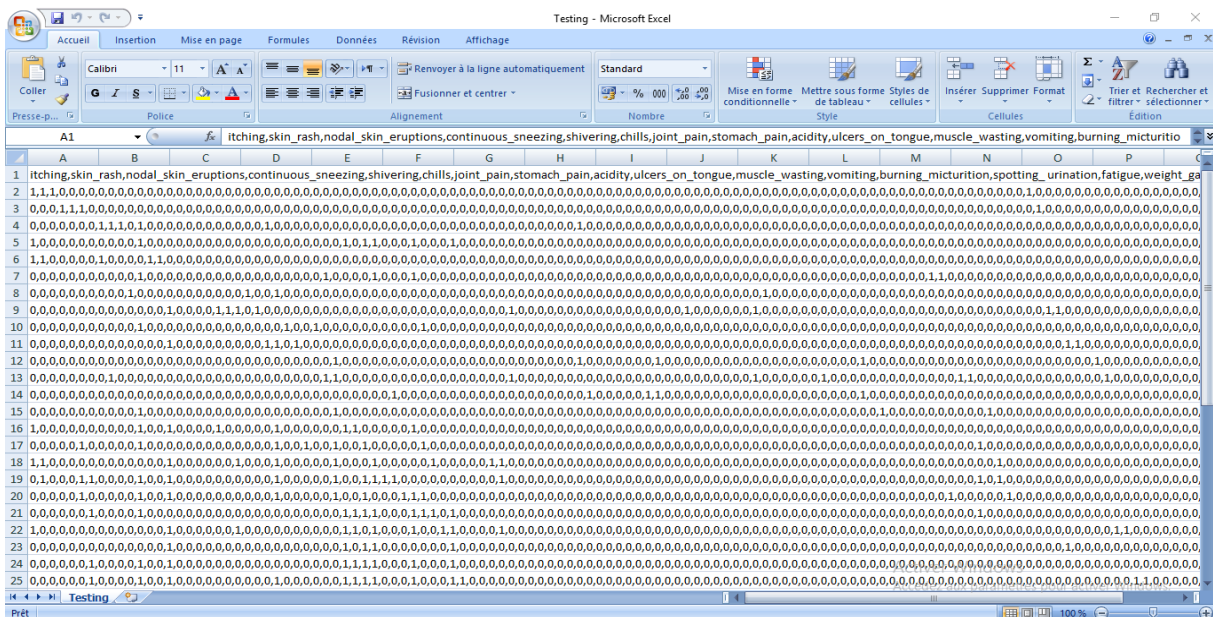


Figure [4.2]: Ensemble de données de test (Testing Data)

4. Interfaces du système :

Une fois l'utilisateur est sur le système il se retrouve sur cette interface :

Cette interface se compose d'un champ de saisie de nom d'utilisateur, et 5 champs optionnaire pour choisir les symptomes, ce n'est pas obligatoir de choisir tout les symptomes mais le plus de symptomes choisi, la plus precise est la prediction.

Ensuite on trouve un bouton "Predict" qui fait la fonction de prediction apres la saisi de nom d'utilisateur et les symptomes, et un bouton "Reset" pour actualiser la page si l'utilisateur veut tout annuler .

Disease Prediction - Prediction de maladie

Select Symptoms

Name

Select Symptoms:

None

None

None

None

None

PREDICT

RESET

Prediction will show here

Activier
Accédez a

Figure [4.3] : interface générale du système

Select Symptoms

Name

Select Symptoms :

▾

▾

▾

▾

▾

According to the symptoms you may have 'Varicose veins'

Figure [4.4] : Interface de résultat de prédiction (1)

Select Symptoms

Name

Select Symptoms :

▾

▾

▾

▾

▾

According to the symptoms you may have 'Urinary tract infection'

Figure [4.5] : Interface de résultat de prédiction (2)

Select Symptoms

Select Symptoms :

Altered Sensorium

Anxiety

Altered Sensorium

Bloody Stool

Bloody Stool

PREDICT

RESET

Prediction will show here

Figure [4.6] : Interface de tentative de prédiction sans saisir le nom.

5. Conclusion

Dans ce chapitre nous avons présenté la manière dont nous avons réalisé ce projet, les différents outils de programmation ainsi que les bibliothèques et les ensembles de données (entraînement, test) utilisés pour notre implémentation, suivie des captures d'interfaces de fonctionnalités du système.

Conclusion générale

Pour conclure, on ajoute par dire que ce projet de prédiction de maladies base sur l'apprentissage automatique est très utile dans la vie quotidienne de chacun de nous, et principalement plus important dans le secteur de sante.

De nos jour l'industrie de la sante joue un rôle majeur dans la guérison des maladies, donc ce système ajoute une sorte d'utilité pour le secteur sante, ainsi que pour l'utilisateur pour moins de déplacement et moins de cout.

La prédiction de maladies consiste à fournir une prédiction de différentes maladies générale parfois ignorées, qui peuvent se développer en problèmes sérieux de sante.

Au terme de ce mémoire de master , nous pouvons résumer les choses en disant que le travail que nous avons mené sur le système de prédiction de maladies, nous a permis dans une première phase d'appréhender l'étendue et la complexité des concepts et des méthodes afférents à ce domaine.

Fort heureusement pour l'élaboration du système, nous ne sommes pas partis de zéro puisqu'il existe déjà des outils d'apprentissage automatique qui nous ont inspiré et que nous avons avantageusement utilisé.

Toutefois, cet effort ne peut être considéré que comme un début pour un travail plus profond qui aurait comme objectif une optimisation plus poussée sur les détails de maladies et les dernières pathologies révélées, une interface utilisateur plus interactive, et rendre ce système une application mobile simple et facile a utilise.

Bibliographie

- [1] : <https://techtomed.com/lintelligence-artificielle-au-service-du-medicament/>.
- [2] : Tie J, Wang Y, Tomasetti C, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* 2016; 8 : 346ra92. [CrossRef] [PubMed] [Google Scholar].
- [3] : Ghasemi M, Nabipour I, Omrani A, et al. Precision medicine and molecular imaging: new targeted approaches toward cancer therapeutic and diagnosis. *Am J Nucl Med Mol Imaging* 2016 ; 6 : 310–327. [PubMed] [Google Scholar]
- [4] : Martin SD, Coukos G, Holt RA, Nelson BH. Targeting the undruggable: immunotherapy meets personalized oncology in the genomic era. *Ann Oncol* 2015; 26 : 2367–2374. [PubMed] [Google Scholar]
- [5] : Mohebian MR, Marateb HR, Mansourian M, et al. A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Comput Struct Biotechnol J* 2017; 15 : 75–85. [CrossRef] [PubMed] [Google Scholar]
- [6] <https://www.ibm.com/fr-fr/topics/artificial-intelligence-healthcare>
- [7] <https://machinelearnia.com/apprentissage-supervise-4-etapes/>
- [8] <https://siecledigital.fr/2018/08/20/histoire-intelligence-artificielle/>
- [9] <https://slideplayer.fr/slide/16258499/>
- [10] <https://www.pierre-giraud.com/html-css-apprendre-coder-cours/definition-utilisation/>
- [11] <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/>
- [12] <https://www.kaherecode.com/tutorial/demarrer-avec-flask-un-micro-framework-python>
- [13] <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/jupyter-notebook/>
- [14] <https://datascientest.com/intelligence-artificielle-definition>
- [15] <https://datascientest.com/knn>