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

Centre Universitaire Belhadj Bouchaib d'Aïn-Témouchent

Institut des Sciences
Département des Mathématiques et de l'Informatique

Mémoire

En vue de l'obtention du Diplôme de Master en Informatique

Option :

Réseaux et Ingénierie des Données (RID)

Présenté par :
Mme.BOUDJEMAI Khadidja

Mme.BOUKRAA Kawther

L'utilisation de service web et des réseaux de neurones pour le diagnostic médical à distance

Encadrant :
Dr BENDIABDALLAH Mohammed Hakim
Maitre de conférence "B" à C.U.B.B.A.T.
Soutenu le : 23/09/2020

Devant le jury composé de :

Présidente :	*Mme SAIDI Samira (M.A.A)*	C.U.B.B.A.T.
Examinatrice :	*Mlle BERRAKEM Fatima Zahra (M.A.A)*	C.U.B.B.A.T.

Année universitaire 2019/2020

Remerciements

Nous remercions en premier lieu Dieu qui nous a donné la chance, la santé, le courage et la volonté pour réussir dans ce travail.

Toute gratitude s'adresse aussi à notre rapporteur, Dr Mohamed Hakim BENDIABDALLAH Maitre de conférence au CUBBAT, pour son appui, son aide, ses orientations, ses conseils et surtout pour sa patience avec nous tout au long de l'élaboration de ce mémoire.

Nous tenons à remercier aussi les membres de jury pour l'intérêt qu'ils ont porté à nous en acceptant d'évaluer ce travail.

BOUDJEMAI & BOUKRAA

Dédicaces

Je dédie ce travail accompagné d'un profond amour :
Tout d'abord, à mon cher père support de ma vie. Tu as toujours été à mes côtés pour me soutenir et m'encourager.
A ma chère mère, flamme de ma vie et source d'amour, de force et d'affection. Merci pour vos efforts qui m'ont guidé tout au long de mon parcours.

BOUDJEMAI Khadidja

Je dédie ce mémoire
Á mon cher père qui n'a jamais cessé de formuler des prières pour moi, de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs.
À ma chère mère source de vie, qui a toujours guidé mes pas vers le bon chemin, à celle qui a tout fait pour que je réussisse.

BOUKRAA Kawther

Table des matières

Table des figures

Liste des tableaux

Liste des acronymes

IA : Intelligence Artificielle
ML : Machine Learning
KNN : k Nearest Neighbors
SVM : Support Vector Machine
AR : Arbre de Décision
RNA : Réseau de Neurone Artificiel
PMC : Perceptron Multi Couche
IMC : Indice de Masse Corporelle
Gros : Grossesses
Glu : Glucoses
PA : Pression Artérielle
EP : épaisseur de la peau
INS : Insuline
DPF : Diabéte Pedigree Fonction
VN : Vrai Négatifs
VP : Vrai Positifs
FN : Faux Négatifs
FP : Faux Positifs
IHM : Interface Homme Machine

Introduction générale

Depuis la venue de l'informatique, l'ensemble des données stockées sous forme numérique ne cesse de croitre de plus en plus rapidement partout dans le monde. Les individus mettent de plus en plus les informations qu'ils possèdent à disposition de tous via le web. De nombreux processus industriels sont également de plus en plus contrôlés par l'informatique. Les résultats d'analyses médicales sont aussi de plus en plus régulièrement conservés pour être analysés, et de nombreuses mesures météorologiques, remplissent aussi d'importances bases de données numériques.

L'intelligence artificielle est un sous-domaine de l'informatique. Ces derniers temps, l'expression «intelligence artificielle » est fréquemment utilisé dans le publique car il s'agit d'un domaine en constante évolution notamment grâce aux progrès des techniques informatiques et entre autres grâce aux capacités toujours plus grandes des machines pour effectuer les calculs.

La machine Learning ou « apprentissage automatique » est un concept qui fait de plus en plus parler de lui dans le monde de l'informatique, et est un sous-domaine de l'intelligence artificielle. Ce terme renvoie à un processus de développement, d'analyse et d'implémentation conduisant à la mise en place de procédés systématiques.Pour faire simple, il s'agit d'une sorte de programme permettant à un ordinateur ou à une machine un apprentissage automatisé, de façon à pouvoir réaliser un certain nombre d'opérations très complexes.

Notre travail s'inscrit dans le cadre de la classification des données numériques en utilisant les techniques de l'apprentissage automatique, et les réseaux de neurones pour développer un service web de diagnostic médical.
Le diagnostic médical est un processus de classification. L'utilisation de l'informatique pour la réalisation de cette classification devient de plus en plus fréquente. Même si la décision de l'expert est le facteur le plus important lors du diagnostic, les systèmes de classification fournissent une aide substantielle, car elles réduisent les erreurs dues à la fatigue et le temps nécessaire pour le diagnostic.

Nous avons organisé la structuration de notre mémoire en trois principaux chapitres :
Chapitre 1 : en premier lieu, nous avons présenté une généralité sur l'intelligence artificielle ainsi que ses axes. Puis, nous avons distingué les différentes approches adoptées pour l'apprentissage automatique et leurs types jusqu' aux algorithmes utilisés pour l'apprentissage supervisé et non supervisé.

Chapitre 2 : dans ce chapitre nous sommes surtout intéressés au réseau de neurones artificiel qui est la méthode choisie dans notre projet, et aussi une description détaillée de la base utilisée (Pima).

Chapitre3 : concerne l'expérimentation, il présente les critères d'évaluation, deux méthodes de sélection des attributs, les expérimentations réalisées, les résultats obtenus avec leur interprétation et une étude comparative avec d'autres types d'algorithmes portant sur le même sujet,en dernier lieu on présente la conception de notre application web ,ainsi que les environnements ulilisés pour la réalisation de cette application.

Enfin, nous terminons par une conclusion générale et quelques percpectives.

Chapitre 1

Etat de l'art sur l'apprentissage automatique

1.1 Introduction

Depuis un demi-siècle, les chercheurs en intelligence artificielle travaillent à programmer des machines capables d'effectuer des tâches qui requièrent de l'intelligence.

Nous citerons l'aide à la décision : l'aide au diagnostic médical, la reconnaissance de formes : la reconnaissance de la parole ou la vision artificielle, la conduite de robots, l'exploration de grandes bases de données on peut dire aussi la fouille de données ou data mining en anglais.

Ce chapitre dresse les notions fondamentales du domaine de classification des données le domaine de notre travail. Il commence par définir l'intelligence artificielle et une distinction entre intelligences artificielles, machine automatique l'apprentissage profond, puis décrire l'apprentissage automatique (machine Learning) pour ensuite aborder ses différents types : supervisé, Semi supervisé et non supervisé, après nous avons défini quelques exemples des algorithmes d'apprentissage automatique.

1.2 Intelligence artificielle

1.2.1 Définition d'intelligence artificielle

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

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 ».[1]

L'intelligence artificielle est une discipline de l'informatique intimement liée à d'autres sciences : les mathématiques, la logique et les statistiques qui lui servent de base théorique, les sciences humaines (sciences cognitives, psychologie, philosophie, linguistique, ...) et la neurobiologie qui aident à reproduire des composantes de l'intelligence humaine par biomimétisme, et enfin, les techniques matérielles qui servent de support physique à l'exécution des logiciels d'IA.[1]

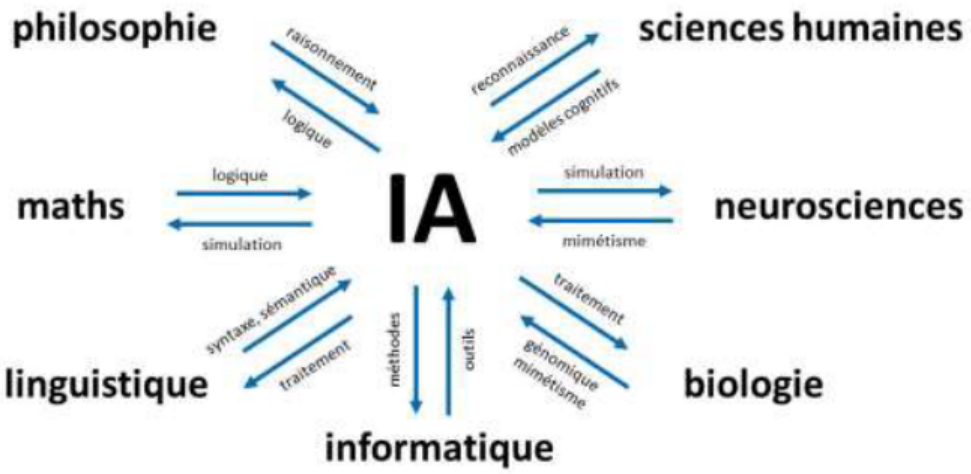

Figure 1.1: Les usages de l'intelligence artificielle [1]

1.2.2 Les axes de l'intelligence artificielle

On parle par l'intelligence artificielle d'une machine qui pense et agir raisonnablement comme l'humain :

1.2.2.1 l'approche humaine

— **La machine doit agir comme les humains :**
Alan Turing c'est le premier qui réfléchit pour créer une machine intelligent. Pour qu'une machine soit considérée intelligente, elle doit passer ce test avec brio. Sans rentrer dans les détails de ce test, ce dernier met en relation un questionneur (humain) à un répondant (humain ou virtuel). Si à la fin du test, le questionneur n'est pas capable de dire si le répondant était un homme ou une machine, alors le test est réussi.[28]

— **La machine doit penser comme les humains :**
Pour que les chercheurs avoir une machine qui pense comme un humain ; Ils doivent avoir premièrement comment l'humain pense et alors ils ont trouvé trois techniques pour tenter de déterminer le fonctionnement de la pensée :
-l'introspection (se saisir de ses propres pensées).
-La psychologie (observer un individu dans ses actions).
-L'imagerie cérébrale.
Une fois que l'on dispose d'une vision assez claire de l'esprit, notamment grâce à toutes les recherches qui ont été faites sur ce sujet, cette théorie représente le cerveau comme un programme informatique. Il suffirait donc de trouver le code de notre cerveau pour pouvoir l'appliquer à un ordinateur.[28]

1.2.2.2 L'approche rationnelle

— **La machine doit agir rationnellement :**
Pour qu'une machine être rationnel, elle doit théoriquement être capable de :
1. fonctionner de façon autonome.
2. Percevoir son environnement.
3. Persister pendant une période prolongée.
4. S'adapter aux changements.
5. Poursuivre des objectifs.

— **La machine doit penser rationnellement :**
Cette approche porte le nom de « loi de la pensée ». Aristote fut l'un des premiers à partir du constat que certaines choses sont toujours vraies. Il a donc voulu codifier le « bien penser », soit les procédés de raisonnement irréfutables. Je cite son exemple : « Socrate et homme, tous les hommes sont mortels, Socrate est donc mortel ». Ce système de pensée est dit « logique ». Et ces lois de la pensée étaient censées régir tout le fonctionnement de l'esprit humain. À l'aide de ce constat sont nés les premiers programmes informatiques capables à résoudre des problèmes logiques. Ces systèmes sont déjà considérés comme « intelligents » car ils permettent d'assister l'Homme dans des taches qui requièrent un raisonnement rationnel.[28]

1.2.3 Distinction entre intelligences artificielles, machine automatique l'apprentissage profond

Il y a une confusion fréquente dans le débat public entre « intelligences artificielles », apprentissage automatique (machine Learning) et apprentissage profond (Deep Learning). Pourtant, ces notions ne sont pas équivalentes, mais sont imbriquées.[2]
-l'intelligence artificielle englobe la machine automatique, qui lui-même englobe l'apprentissage profond.

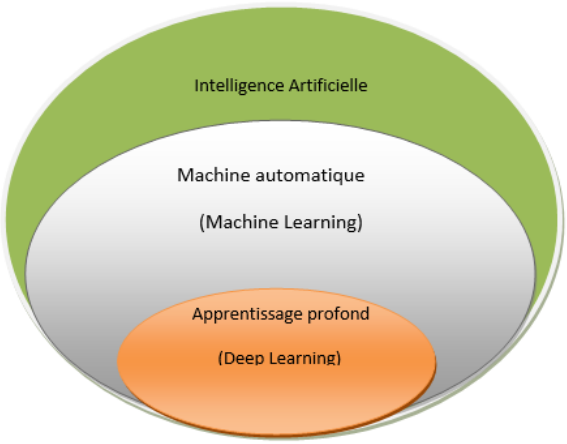

Figure 1.2: Positionnement de l'IA d'un point de vue global par rapport au machine Learning et au Deep learning

1.3 Apprentissage automatique (machine Learning)

1.3.1 Définition

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle, où le terme fait référence à la capacité des systèmes informatiques à trouver indépendamment des solutions aux problèmes en reconnaissant les modèles dans les bases de données. En d'autres termes : la Machine Learning (ML) permet aux systèmes informatiques de reconnaître des modèles sur la base d'algorithmes et d'ensemble de données existantes et de développer des concepts de solutions adéquates. Par conséquent, dans la Machine Learning, la connaissance artificielle est générée sur la base de l'expérience.[3]

1.4 Fonctionnement de l'apprentissage automatique

D'une certaine manière, la Machine Learning fonctionne de manière similaire à l'apprentissage humain. Par exemple, si un enfant voit des images avec des objets spécifiques sur lui, il peut apprendre à les identifier et à les différencier. L'apprentissage automatique fonctionne de la même manière : grâce à la saisie de données et à certaines commandes, l'ordinateur est en mesure «d'apprendre» à identifier certains objets (personnes, objets, etc.) et à les distinguer. À cet effet, le logiciel est fourni avec des données et formé. Par exemple, le programmeur peut dire au système qu'un objet particulier est un être humain (= "humain") et qu'un autre objet n'est pas un être humain (= "pas d'humain"). Le logiciel reçoit une rétroaction continue du programmeur. Ces signaux de rétroaction sont utilisés par l'algorithme pour adapter et optimiser le modèle. Avec chaque nouvel ensemble de données introduit dans le système.[3]

1.5 Types d'apprentissage automatique

Bien qu'un modèle d'apprentissage automatique puisse appliquer un mélange de différentes techniques, les méthodes d'apprentissage peuvent généralement être classées en trois types généraux :

1.5.1 Apprentissage supervisé

Avec la classification qui permet de labelliser des objets comme des images et la régression qui permet de réaliser des prévisions sur des valeurs numériques. L'apprentissage est supervisé car il exploite des bases de données d'entraînement qui contiennent des labels ou des données contenant les réponses aux questions que l'on se pose. En gros, le système exploite des exemples et acquiert la capacité à les généraliser ensuite sur de nouvelles données de production.[1]
Il est divisé en deux méthodes :

1.5.1.1 Classification

Il s'agit de pouvoir associer une donnée complexe comme une image ou un profil d'utilisateur à une classe d'objets, les différentes classes possibles étant fournies à priori par le concepteur.
La classification utilise un jeu de données d'entraînement associé à des descriptifs (les classes) pour la détermination d'un modèle. Cela génère un modèle qui permet de prédire la classe d'une nouvelle donnée fournie en entrée. Dans les exemples classiques, nous avons la reconnaissance d'un simple

chiffre dans une image, l'appartenance d'un client à un segment de clients ou pouvant faire partie d'une typologie particulière de clients (mécontents, pouvant se désabonner à un service, etc.) ou la détection d'un virus en fonction du comportement ou de caractéristiques d'un logiciel.[4]

1.5.1.2 Régression

La régression permet de prédire une valeur numérique y en fonction d'une valeur x à partir d'un jeu d'entraînement constitué de paires de données (x, y).
On peut par exemple prédire la valeur d'un bien immobilier ou d'une société en fonction de divers paramètres les décrivant.
Le schéma ci-dessous qui illustre ce concept utilise uniquement une donnée en entrée et une en sortie. Dans la pratique, les régressions utilisent plusieurs paramètres en entrée.

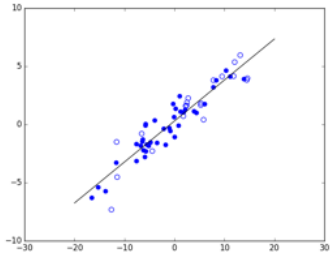

Figure 1.3: Régression linéaire [29]

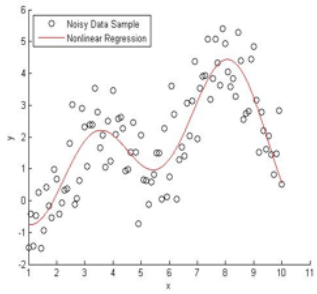

Figure 1.4: Régression non linéaire [30]

Les jeux de données en entrée comprennent plusieurs variables (x, y, z, ...). Il existe différentes formes de régression, notamment linéaire et non linéaire. S'y ajoutent aussi les notions d'over fishing et d'underfitting, qui décrit les méthodes de régression qui suivent plus ou moins de près les variations observées. Il faut éviter les deux et trouver le juste milieu ! C'est le travail des datèrent scientistes.[4]

1.5.2 Apprentissage non supervisé

Avec le clustering et la réduction de dimensions. Il exploite des bases de données non labellisées. Ce n'est pas un équivalent fonctionnel de l'apprentissage supervisé qui serait automatique. Ses fonctions sont différentes. Le clustering permet d'isoler des segments de données spatialement séparés entre eux, mais sans que le système donne un nom ou une explication de ces clusters. La réduction de dimensions vise à réduire la dimension de l'espace des données, en choisissant les dimensions les plus pertinentes. Du fait de l'arrivée des biges data, la dimension des données a explosé et les recherches sur les techniques des dimensions les plus pertinentes sont très actives.[1]

parmi les méthodes d'apprentissage non supervisé :

1.5.2.1 Clustering

Le clustering ou la segmentation automatique est une méthode d'apprentissage non supervisé qui permet à partir d'un jeu de données non labellisé d'identifier des groupes de données proches les unes des autres, les clusters de données.
La technique la plus répandue est l'algorithme des k -Moyennes (k-means).

1.5.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé situé quelque part entre l'apprentissage supervisé et l'apprentissage non supervisé, car il utilise des données étiquetées et non étiquetées pour la formation généralement une petite quantité de données étiquetées et une grande quantité de données non étiquetées.
Les systèmes qui utilisent cette méthode sont capables d'améliorer considérablement la précision d'apprentissage. Habituellement, l'apprentissage semi-supervisé est choisi lorsque les données labellisées acquises nécessitent des ressources compétentes et pertinentes pour les former / en tirer des enseignements. Sinon, l'acquisition de données sans étiquette ne nécessite généralement pas de ressources supplémentaires.[31]

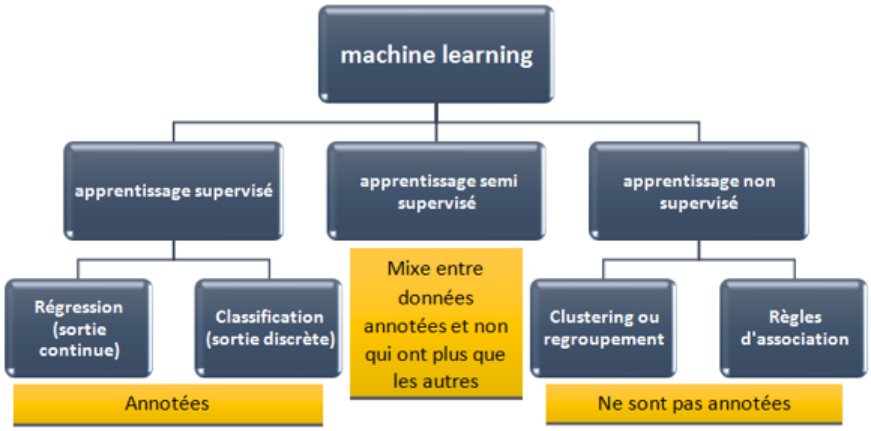

Figure 1.5: Schéma illustratif des types d'apprentissage automatique

1.6 Quelques exemples d'algorithmes d'apprentissage automatique

1.6.1 Arbre de décision

a) **Définition :**
L' arbre de décision est l'outil le plus puissant et le plus populaire pour la classification et la prédiction. Un arbre de décision est un organigramme semblable à une structure arborescente, où chaque nœud interne(nœud de décision) désigne un test sur un attribut, chaque branche représente un résultat du test et chaque nœud feuille (nœud terminal) est repéré par sa position (liste des numéros des arcs qui permettent d'y accéder en partant de la racine),et étiquetées par une classe.[5]

Lors de l'apprentissage d'un arbre, les données source sont divisées en sous-ensembles en fonction d'un test de valeur d'attribut, qui est répété récursivement sur chacun des sous-ensembles dérivés. Une fois que le sous-ensemble d'un nœud a la valeur équivalente à sa valeur cible, le processus sera terminé.[5]

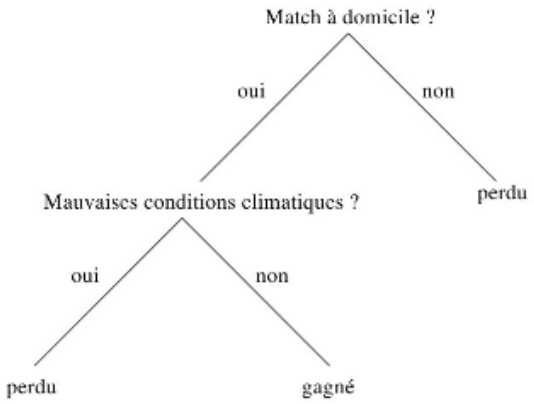

Figure 1.6: Exemple d'un arbre de décision [6]

b) **Avantages :**

- Décisions aisément interprétables.
- classification très rapide.
- Peu de traitements sur les données.

c) **Inconvénients :**

- On peut se trouver avec des arbres de décision finale très complexe.
- Certains concepts sont difficiles à exprimer à l'aide d'arbres de décision (comme XOR ou la parité).

1.6.2 K plus proches voisin

L'algorithme KNN (k-Nearest Neighbors) figure parmi les plus simples algorithmes d'apprentissage artificiel. Dans un contexte de classification d'une nouvelle observation x, l'idée fondatrice simple est de faire voter les plus proches voisins de cette observation. La classe de x est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation x. Donc la méthode du plus proche voisin est une méthode non paramétrique où une nouvelle observation est classée dans la classe d'appartenance de l'observation de l'échantillon d'apprentissage qui lui est la plus proche, au regard des Covariables utilisées. La détermination de leur similarité est basée sur des mesures de distance.[6]

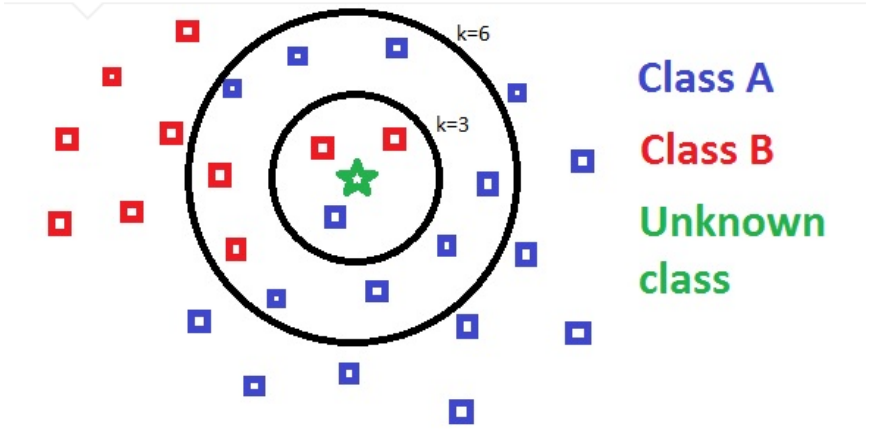

Figure 1.7: Illustration du K plus proches voisin[32]

1.6.3 Machine à vecteurs de support (SVM)

Les machines à vecteurs de supports ou séparateurs à vaste marge (en anglais support Victor Machine, SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de classification. les SVM sont une généralisation des classifiés linéaires.

Les SVM ont été développés dans les années 1990 à partir des considérations théoriques de Vladimir Vapnik sur le développement d'une théorie statistique de l'apprentissage : la théorie de Vapnik Chervonenkis. Les SVM ont rapidement été adoptés pour leur capacité à travailler avec des données de grandes dimensions, le faible nombre d'hypers paramètres, leurs garanties théoriques, et leurs bons résultats en pratique.

Les SVM ont été appliqués à de très nombreux domaines (bio-informatique, recherche d'informations, vision par ordinateur, finance...). Selon les données, la performance des chapitres est : La recherche et la classification d'image : Notions, Méthodes 17 machines à vecteurs de support sont du même ordre, ou mêmes supérieures, à celle d'un réseau de neurones ou d'un modèle de mixture gaussienne. Hyperplan qui est le lieu des points x satisfaisant +b=0. En orientant l'hyperplan, la règle de décision correspond à observer de quel côté de l'hyperplan se trouve l'exemple x. On voit que le vecteur w définit la pente de l'hyperplan (w est perpendiculaire à l'hyperplan). Le terme b quant à lui permet de translater l'hyperplan parallèlement à lui-même. ; Décision h(s). La classe de tous les hyperplans qui en découle sera notée H.[6]

1.6.4 Réseau de neurones artificiels

Un réseau de neurones artificiels, ou réseaux connexionnistes, est un système informatique matériel ou logiciel dont le fonctionnement est calqué sur celui des neurones du cerveau humain. Les premiers travaux datent de 1943 par deux chercheurs de l'Université de Chicago.

Les réseaux de neurones artificiels sont fondés sur des modèles qui tentent de mimer les cellules du cerveau humain et leurs interconnexions. Le but, d'un point de vue global, est d'exécuter des calculs complexes et de trouver, par apprentissage, une relation non linéaire entre des données numériques et des paramètres.

1.7 Application et exemples d'apprentissage automatique

L'apprentissage automatique est un domaine d'étude et une approche de la résolution de problèmes. Et il existe de nombreuses applications différentes auxquelles les méthodes d'apprentissage automatique peuvent être appliquées. Voici quelques-unes des nombreuses applications des stratégies et méthodes d'apprentissage automatique :

1.7.1 Traitement du langage naturel

Le traitement du langage nature est un domaine de la linguistique, de l'informatique qui s'intéresse à l'interaction entre ordinateur et langage humains (naturels). Il fait partie des techniques d'intelligence artificielle. Les algorithmes d'intelligence artificielle ont pour rôle d'identifier et d'extraire les règles du langage naturel, afin de convertir les données de langage non structuré sous une forme que les ordinateurs pourront comprendre. La plupart des techniques de traitement naturel du langage reposent sur l'apprentissage profond.[33]

1.7.2 Moteur de recherche

Un moteur de recherche est une application informatique permettant de rechercher des ressources, des contenues, des documents (page web, d'images, de vidéos, d'actualités, de fichiers, etc...), à partir de mots clés.

La recherche sur le Web repose sur l'utilisation du Machine Learning en l'utilisant pour mieux comprendre les requêtes des utilisateurs et améliorer les résultats de recherche . Il existe différents moteurs de recherche. Le plus connu, et celui que vous utilisez sans doute : c'est Google. Des entreprises comme Google peuvent améliorer leurs résultats de recherche et comprendre quels sont les meilleurs résultats pour une requête donnée. Des suggestions de recherche et des corrections d'orthographe sont également générées en utilisant des tactiques d'apprentissage automatique sur les requêtes agrégées de tous les utilisateurs.[33]

1.7.3 Bio-informatique et diagnostic médical

Ces dernières années, l'émergence de nouvelles technologies en biologie ainsi que les avancées en informatique ont augmenté non seulement la quantité des données biologiques mais aussi leur complexité.

Les scientifiques font alors appel aux technologies de l'informatique et déployés des techniques d'apprentissage automatique qui leur permettent de transformer ces données en information et de résoudre ainsi des problèmes biologiques et médicaux afin de classer et de mieux comprendre diverses maladies. Ces approches devraient également aider à diagnostiquer la maladie en identifiant les segments de la population qui sont les plus à risque de certaines maladies.[33]

1.7.4 Traitement d'images et reconnaissance de formes

L'utilisation d'ordinateurs pour identifier des formes (ou parfois reconnaissance de motifs) et identifier des fichiers multimédias (images, des vidéos, etc.)Est beaucoup moins pratique sans techniques d'apprentissage automatique. Écrire des programmes pour identifier des motifs informatiques à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à ce motif.

Les algorithmes de reconnaissance d'images (classificateurs d'images) peuvent être formés pour classer les images en fonction de leur contenu. Ces algorithmes sont formés en traitant de nombreux exemples d'images qui ont déjà été classés. En utilisant les similitudes et les différences d'images qu'ils ont déjà traitées, ces programmes s'améliorent en mettant à jour leurs modèles chaque fois qu'ils traitent une nouvelle image. Cette forme d'apprentissage automatique utilisée dans le traitement d'images est généralement effectuée à l'aide d'un réseau de neurones artificiels et est connue sous le nom d'apprentissage automatique.[33]

1.8 Conclusion

Ce qu'il faut retenir de tout ce qui précède, c'est tout simplement que l'intelligence artificielle semble promise à un bel avenir, au vu de la portée technologique de la machine Learning. La machine Learning offre un certain nombre de méthodes statistiques avancées pour traiter des tâches de régression et de classification avec plusieurs variables dépendantes et indépendantes. Citons la méthode des séparateurs à Vaste Marge (SVM - Support Vecteur Machines), et K. plus Proches Voisins pour des problèmes de régression et de classification, et les méthodes arbre de décision et réseau de neurones pour des problèmes de classification. Vous trouverez une présentation détaillée de la méthode choisie pour réaliser notre travail (réseau de neurones artificiels (RNA)) dans le chapitre suivant.

Chapitre 2

Réseau neuronal artificiel (RNA) et Base de données utilisées

2.1 Introduction

Durant ces dernières années les réseaux de neurones artificiels se sont imposés dans plusieurs domaines comme des classificateurs ou bien des détecteurs.

Ce chapitre a pour objectif de présenter le réseau de neurones artificiels qu'on a choisi pour utiliser dans notre application ; qui constituent une approche permettant d'aborder sous des angles nouveaux les problèmes de perception, de mémoire, d'apprentissage et de raisonnement, ils se révèlent aussi des alternatives pour éviter les limitations des méthodes classiques grâce à leur traitement parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de solutionner des problèmes complexes.

On va présenter aussi dans cette partie la base qu'on a utilisée pour notre apprentissage automatique c'est une base qui compose d'ensemble des malades diabétiques s'appelle la base « Pima ».

2.2 Définition

Comme leur nom l'indique, les réseaux de neurones sont inspirés de l'architecture du cerveau (neurones biologiques), étant organisés en couches de neurones connectées entre elles. Un réseau de neurones est un ensemble de méthodes d'analyse et de traitements des données permettant de construire un modèle de comportement à partir de données qui sont des exemples de ce comportement. Un réseau de neurones est constitué d'un graphe pondéré orienté dont les nœuds

2.2.1 Neurone biologique

Un neurone est une cellule du système nerveux capable de communiquer et de traiter des informations. On le trouve dans le cerveau mais aussi dans la moelle épinière et les nerfs optiques. De façon très réductrice, un neurone biologique est une cellule qui se caractérise par :[8]

- **corps cellulaire (soma) :** Cette partie centrale a pour rôle de contrôler la réaction de la cellule, en fonction des informations reçues en entrée.
- **dendrites :** Ce sont elles qui transmettent les signaux depuis l'extérieur vers le corps cellulaire.
- **Axone :** permets de transmettre les signaux électriques, et ainsi de véhiculer le message en sortie.
- **Synapses :** assurent la communication de l'information aux autres neurones, ainsi qu'aux fibres musculaires (pour d'éventuelles réponses nerveuses).

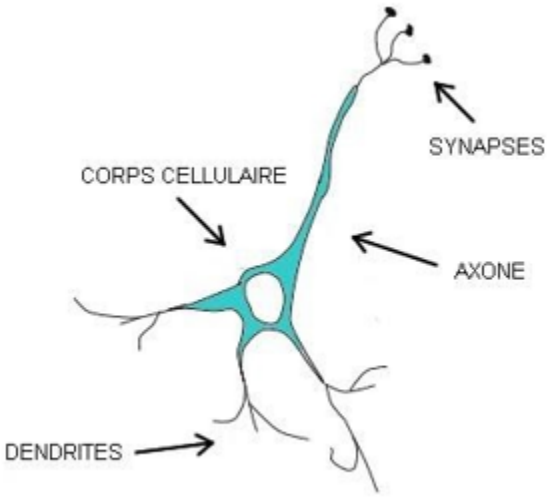

Figure 2.1: Schéma d'un neurone biologique.[34]

2.2.2 Neurone formel

Un neurone formel est une modélisation mathématique et informatique qui reprend les principes du fonctionnement du neurone biologique ou :

- Les synapses sont modélisées par des poids
- Le soma ou corps cellulaire est modélisé par la fonction de transfert, appelé aussi fonction d'activation.
- L'axone est modélisée par l'élément de sortie.

Un neurone formel, est une fonction algébrique non linéaire et bornée, peut être caractérisé par :

a) Plusieurs entrées (x1, x2,. . . , xi,. . . , xn), qui peuvent être les entrées du réseau ou les sorties d'autres neurones du même réseau.

b) Chaque entrée associée à un poids ou coefficient (w1, w2,. . . .... Wi, . . . , wn).

c) La fonction de combinaison : dans le modèle initial, il s'agit simplement d'une somme pondérée des valeurs en entrée :

$$\sum wi.xi + b \tag{2.1}$$

- ou b désigne le seuil d'activation .

d) La fonction d'activation, ou d'état f, définissant l'état interne du neurone en fonction de son entrée totale. Cette fonction peut prendre plusieurs formes :

- linéaire g est la fonction identité.
- *seuil* $g(x) = 1[0, +\infty[(x)$
- *sigmode* $g(x) = 1/(1 + \exp^x)$
- *ReLU* $g(x) = max(0, x)(rectifiedlinearunit)$
- *Softmax* $g(x) = \frac{\exp^x}{\sum_{k=1}^{K} \exp^x K}$
- .....

e) Sa fonction de sortie calculant la sortie du neurone en fonction de son état d'activation.

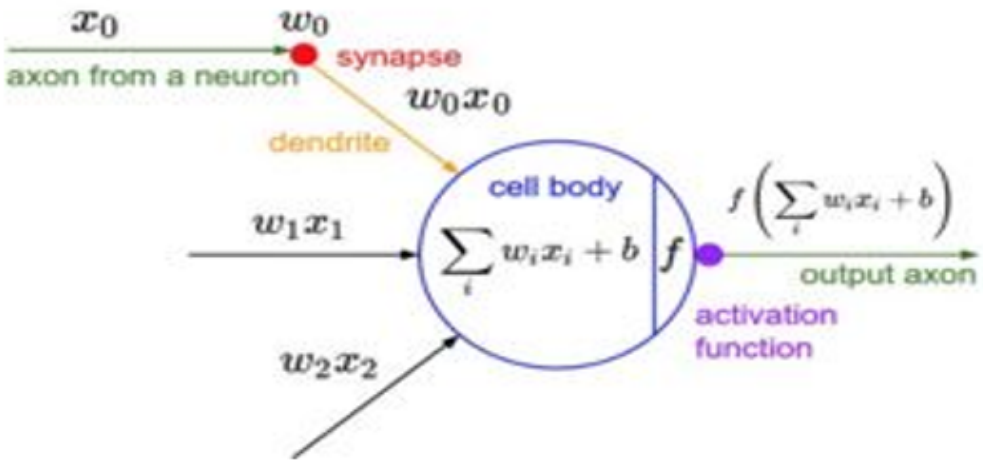

Figure 2.2: Schéma d'un neurone formel.[35]

2.3 Architecture des réseaux de neurones

La plupart des réseaux de neurones sont construits de trois couches successives : une couche d'entrée, une couche cachée, et une couche de sortie .Toutefois, il peut y avoir 0 à N couche cachées.

-Selon la topologie de connexion des neurones, on peut les classer en deux grandes catégories :

2.3.1 Réseau de neurones monocouche(Perceptron)

Dans ce type de réseau, il y a une seule couche cachée, qui relie les cellules d'association (couche d'entrée) aux cellules de décision (couche de sortie). C'est la seule couche déconnexion modifiable. Les neurones de la couche d'entrée d'un réseau monocouche (perceptron) effectuent seulement un prétraitement et la classification effective est effectuée par les neurones de la couche de sortie. Ce réseau offre une grande convergence vers la solution du problème, malheureusement

sa stratégie d'apprentissage n'offre que des séparations linéaires, limitées à la seule classe de problèmes linéairement séparables.[9]

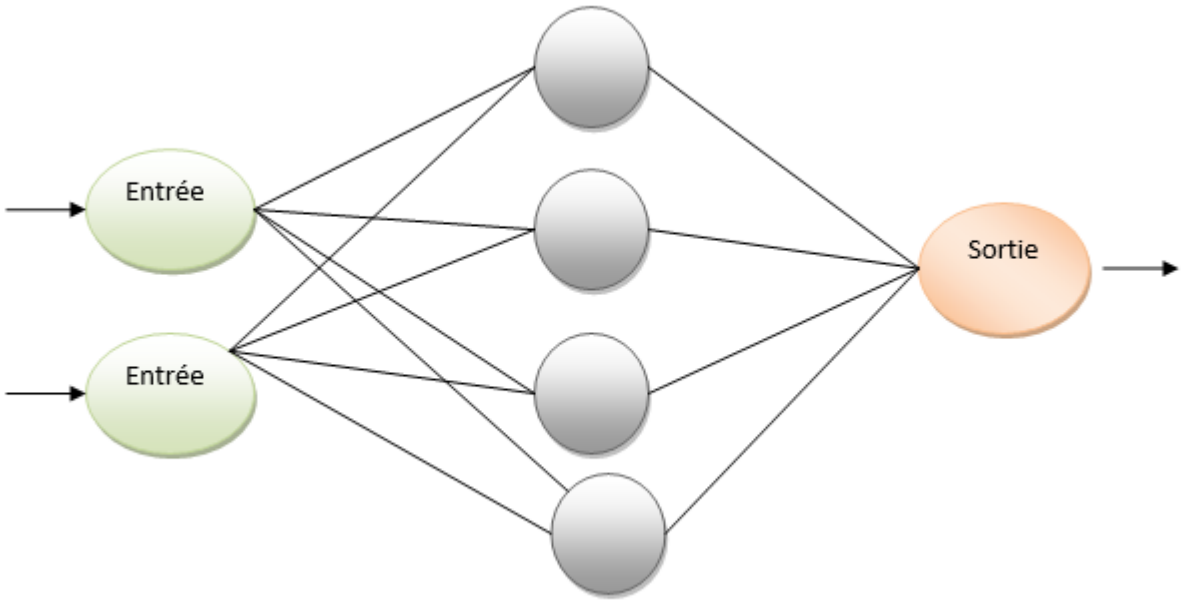

Figure 2.3: Réseau de neurones monocouche.

2.3.2 Réseau de neurones multicouche (PMC : Perceptron Multi Couche)

C'est le réseau de neurones statique le plus utilisé. Les neurones sont arrangés par couches. Les neurones de la première couche reçoivent le vecteur d'entrée, ils calculent leurs sorties qui sont transmises aux neurones de la seconde couche qui calculent eux-mêmes leurs sorties et ainsi de suite de couches en couche jusqu'à celle de sortie. Chaque neurone dans la couche cachée est connecté à tous les neurones de la couche précédente et de la couche suivante, et il n'y a pas de connexions entre les cellules d'une même couche.[10]

Il peut résoudre des problèmes non linéairement séparables et il suit un apprentissage supervisé avec la règle de correction de l'erreur.

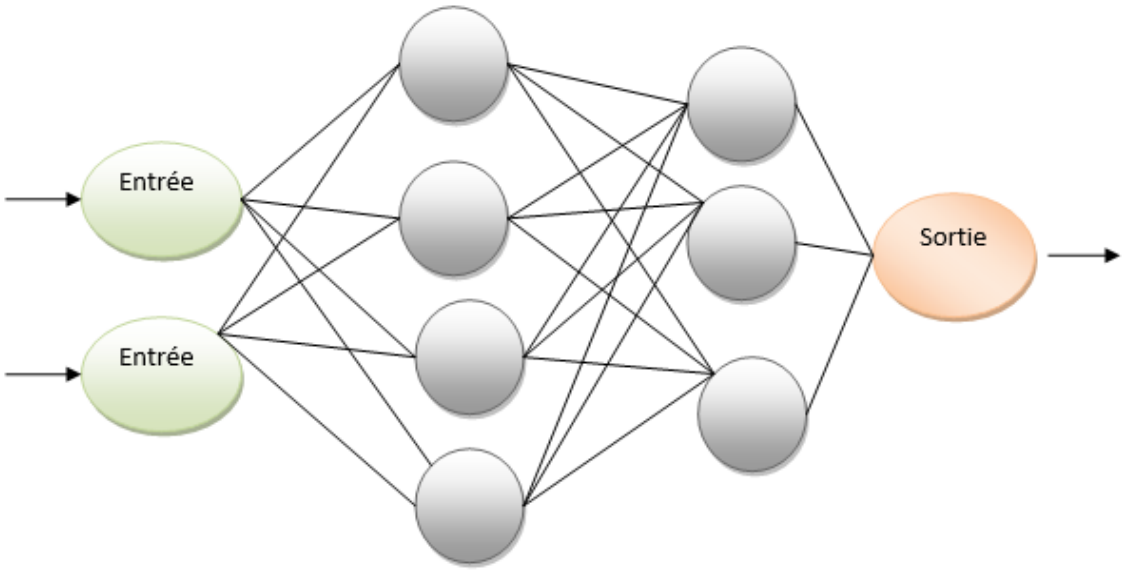

Figure 2.4: Réseau de neurones multicouche.

2.4 Dropout dans les réseaux de neurones

Le dropout est une méthode qui consiste à retirer une partie des neurones du réseau qui sont sélectionnés en utilisant des critères de sélection simples lors de l'entrainement. Le dropout s'attaque au problème de coadaptation des neurones dans le réseau. La coadaptation est une situation où plusieurs neurones d'une même couche sont utilisés pour modéliser une seule information .[11]

Les coadaptations dites complexes apprises par un réseau ne sont pas toujours nécessaires et introduisent deux problèmes :

1) La diminution des capacités de modélisation du réseau. Si plusieurs neurones modélisent la même information, alors ils sont perdus pour en modéliser de nouvelles.
2) Une tendance au surapprentissage.

Le dropout est donc une solution qui permet d'améliorer les résultats et la performance des réseaux de neurones.

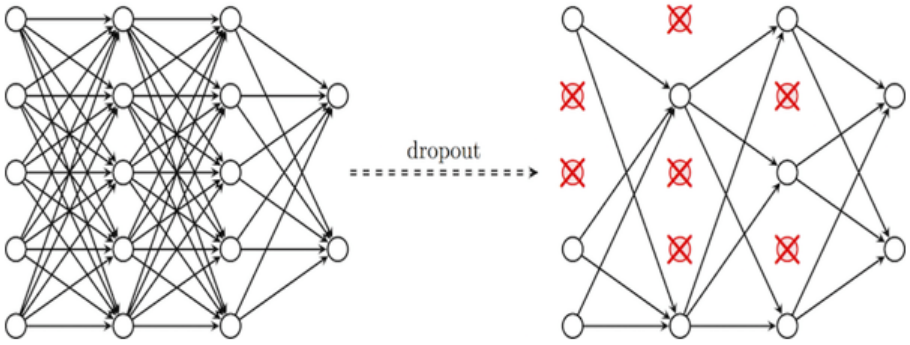

Figure 2.5: Schéma de l'optimisation d'un réseau de neurones avec dropout.[12]

2.5 Base de données sur le diabète des Indiens Pima

2.5.1 Définition

Dans cette mémoire nous utilisons la base de données médicale réelle « Indians Diabètes Pima », cet ensemble de données provient à l'origine de l'Institut national du diabète et des maladies digestives et rénales qui réalise une étude sur 786(instances) femmes Indiennes d'au moins 21 ans d'origine indienne Pima, l'objectif de l'ensemble de données est de prédire par diagnostic si un patient souffre ou non de diabète.[15]

Le diagnostic est une valeur binaire variable «classe» qui permet de savoir si le patient montre des signes de diabète selon les critères de l'Organisation Mondiale de la Santé. Les huit descripteurs (attributs) cliniques sont :

1) Grossesses : nombre de grossesses.
2) Glucose : concentration plasmatique de glucose à 2 heures dans un test de tolérance au glucose par voie orale (cg/l).
3) Pression artérielle : pression artérielle diastolique (mm Hg).
4) Épaisseur de la peau : épaisseur du pli cutané du triceps (mm).
5) Insuline : insuline sérique 2 heures (mu U / ml).
6) IMC : indice de masse corporelle $(poidsenkg/(hauteurenm)^2)$
7) Diabètes Pedigree Fonction : Fonction pedigree du diabète.
8) Âge : Âge (ans).

La dernière colonne de l'ensemble de données indique si la personne est diabétique (1) ou non diabétique (0) « variable de classe », 268 sur 768 sont diabétiques.

2.5.2 visualisation d'ensemble de données

2.5.2.1 Grossesses

La figure 2.6 illustre que le nombre de grossesses est entre un minimum de 0 grossesse et un maximum de 17 grossesses, on remarque que tant que le nombre de grossesses est petit, plus que le risque de développer un diabète est faible par exemple de 0 à 2 grossesses le nombre des femmes non diabétiques est le plus élevé(250 femmes).

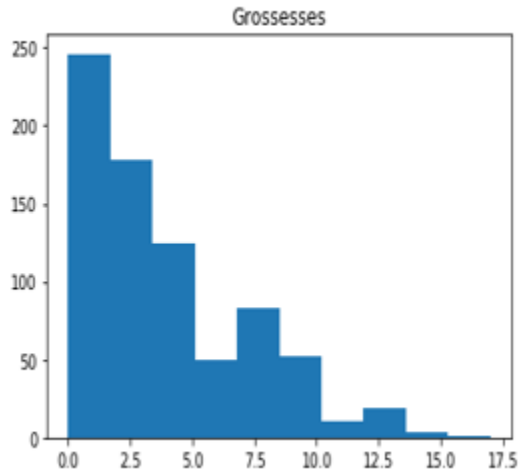

Figure 2.6: Graphe présentatif de l'attribut grossesses

2.5.2.2 Glucose

Le glucose est parmi les facteurs de risque de développer un diabète en analysant ce graphe on observe que la possibilité d'avoir un diabète est élevée lorsque le glucose est moins de 79(Cg/l) et plus de 159 (Cg/l) .On a remarqué aussi qu'il y a 211 personnes qui ne sont pas diabétiques lorsque le glucose est entre 99.5 (Cg/l) et 119.4 (Cg/l)

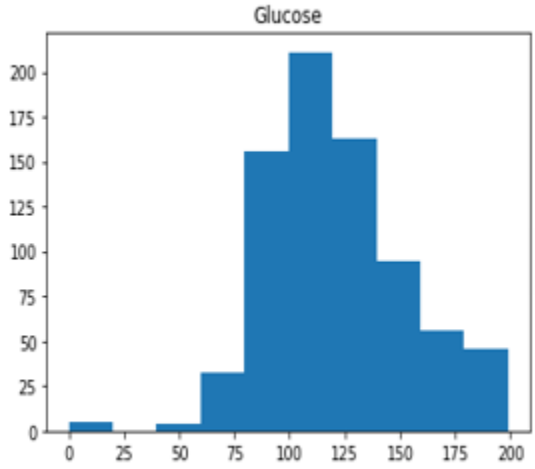

Figure 2.7: Graphe présentatif de l'attribut glucose

2.5.2.3 Pression artérielle

La pression artérielle, correspond à la pression du sang dans les artères de la circulation systémique. D'après ce graphe on a remarqué que lorsque la pression artérielle est entre 61(mm Hg) et 85(mm Hg) le nombre de femmes non diabétiques est élevé et on a observé aussi que les diabétiques ont une pression artérielle plus élevé que les non diabétiques.

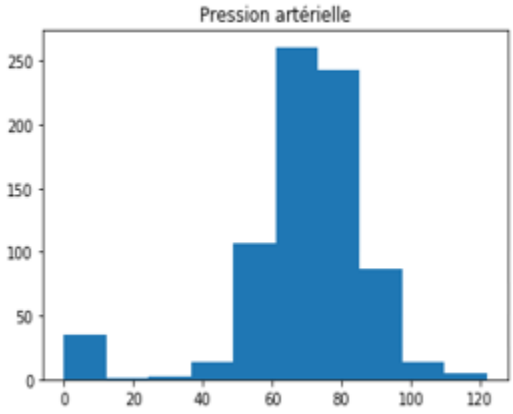

Figure 2.8: Graphe présentatif de l'attribut pression artérielle

2.5.2.4 Épaisseur de la peau

Dans cette figure on remarque que lorsque l'épaisseur de peau est moins la possibilité de n'avoir pas un diabète est élevé donc on peut conclut que l'épaisseur de peau des diabétiques est supérieure à celle des non diabétiques.

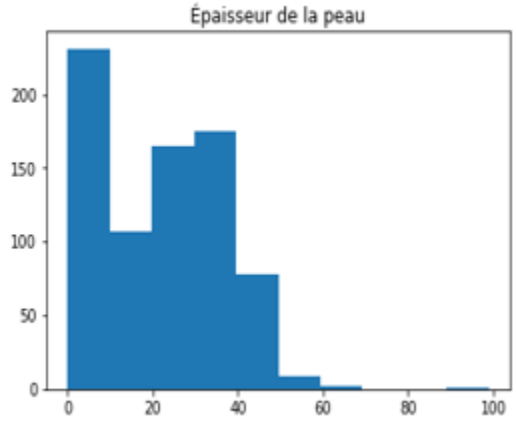

Figure 2.9: Graphe présentatif de l'attribut épaisseur de la peau

2.5.2.5 Insuline

L'insuline est une hormone polypeptidique intervenant dans le cycle du glucose, lorsque sa sécrétion est insuffisante, il y a apparition du diabète. On remarque dans la figure 2.10 que le nombre des diabétiques est élevé (500 femmes) lorsque l'insuline est entre 0 (mu U / ml) et 150 (mu U / ml).

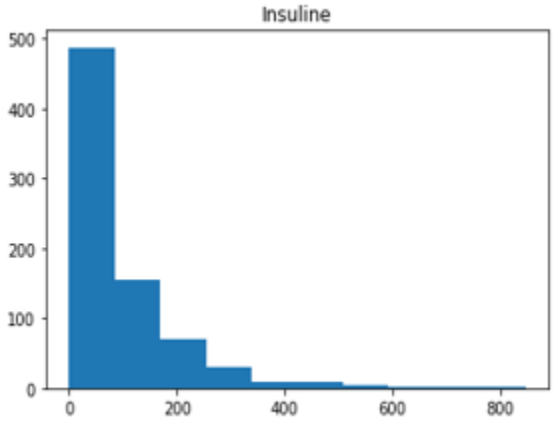

Figure 2.10: Graphe présentatif de l'attribut insuline

2.5.2.6 IMC (Indice de Masse Corporelle)

Indice de masse corporelle permet de savoir si notre poids est idéal, autrement dit s'il est adapté à notre taille, on peut remarquer à partir de la figure 2.11 que si l'IMC est entre 0 et 20 la possibilité d'avoir un diabète est très faible, mais à partir de 20 jusqu'à 40 en observe une augmentation dans le nombre des diabétiques.

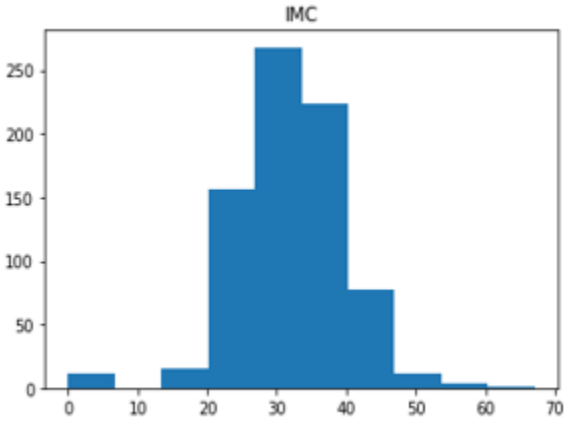

Figure 2.11: Graphe présentatif de l'attribut IMC

2.5.2.7 Fonction pedigree du diabète

Fonction pedigree du diabète a un rôle très important pour avoir un diabète. On a remarqué que la majorité des femmes qui ont une fonction pedigree du diabète entre 0.08 et 0.3 ne sont pas des diabétiques. On peut conclut que les diabétiques semblent avoir une fonction de pedigree plus élevée que les non diabétiques.

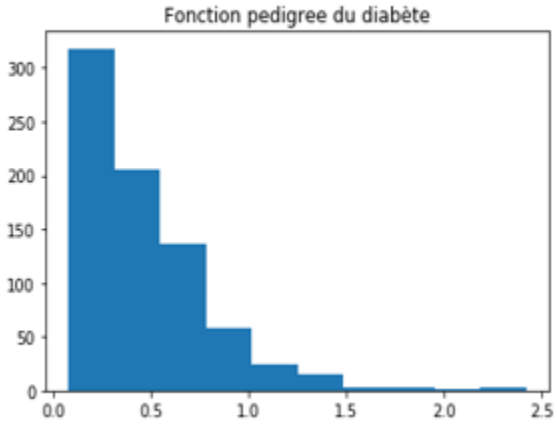

Figure 2.12: Graphe présentatif de l'attribut fonction pedigree du diabète

2.5.2.8 Âge

Le risque de développer un diabète augmente avec l'âge, dans cette base les études sont faites sur une catégorie des femmes d'au moins 21 ans et Max 81 ans .La figure 2.13 nous montre que la catégorie entre 20 et 65 ans a la possibilité d'avoir un diabète avec(8.2 pourcent) c'est pour ça on remarque que 300 des femmes de l'âge entre 21 et 26 ne sont pas diabétiques, et il va démuni le nombre des non diabétique avec l'augmentation de l'âge (à partir de 65 ans jusqu'à 81 ans le risque augmente à "20 pourcent").

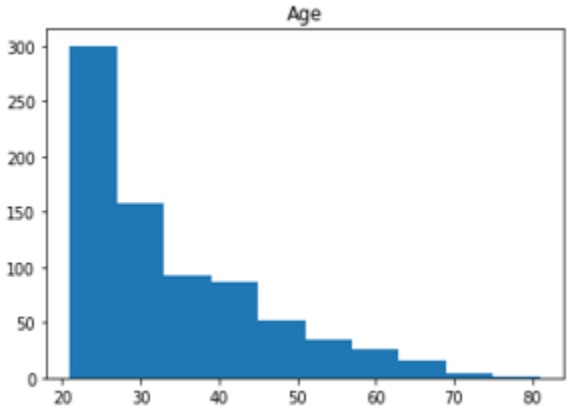

Figure 2.13: Graphe présentatif de l'attribut âge :

2.6 Prétraitement des données

La qualité des données est essentielle pour obtenir des modèles prédictifs performants. Pour éviter de traiter des données erronées et améliorer la performance du modèle, il faut impérativement analyser les données, détecter les anomalies le plus tôt possible et déterminer les étapes de prétraitement.

Le prétraitement des données est une tâche importante qui doit intervenir avant d'utiliser un jeu de données pour la formation de modèles. Les données brutes sont souvent bruyantes, peu fiables et incomplètes. Leur utilisation pour la modélisation peut générer des résultats trompeurs. Ce dernier consiste à traiter les données bruitées, soit en les supprimant, soit en les transformant de manière à en tirer le meilleur profit.

Il existe plusieurs techniques de transformation de données, les plus utilisés sont :

2.6.1 Normalisation

Une normalisation des données d'entrés peut- être appliqué quand les données varient dans des échelles différentes. L'échelle signifie généralement changer la plage des valeurs. La forme de la distribution ne change pas. Pensez à la façon dont un modèle réduit d'un bâtiment a les mêmes proportions que l'original, juste plus petites. C'est pourquoi nous disons qu'il est dessiné à l'échelle. La plage est souvent définie entre 0 et 1.[36]
La transformation se fait grâce à la formule suivante :

$$Xnormalis = \frac{X - Xmin}{Xmax - Xmin} \tag{2.2}$$

Avec :
X min :est la valeur minimale des données.
X max :est la valeur maximale des données.

2.6.2 La standardisation

Cette technique consiste à transformer chaque dimension de données d'entrée de tell sorte aura une valeur moyenne de 0 et un écart type de 1.[36]

calculé comme suit :

$$Xstandr = \frac{X - \mu}{\theta} \tag{2.3}$$

ou :
μ**(mu) :** est la moyenne des données d'entrée x .
θ**(theta) :** est l'ecart-type des données d'entrée x.

2.7 Conclusion

Dans ce chapitre on a discuté sur le réseau d'apprentissage choisit pour notre application qui est le réseau de neurones artificiels ; on a défini aussi son architecture qui est classée en deux grandes catégories : le réseau de neurones monocouche et le réseau de neurones multicouche

Cette partie contient aussi les types du réseau de neurones artificiels, ainsi la base de données médicale «Indians Diabètes Pima» utilisé par notre modèle d'apprentissage.

Chapitre 3

Expérimentation et Résultats

3.1 Introduction

Dans les chapitres précédents nous avons exposé des exemples d'algorithmes d'apprentissage automatique et une description de la base « Pima ».

Nous présentons dans cette partie notre étude comparative, au début nous avons créé deux architectures du réseau neuronal, la première fait entrer les huit paramètres de la base Pima, et la deuxième sélectionne seulement quatre attributs qui sont les plus sensibles pour cette maladie. Après nous avons réalisé une étude comparative entre notre résultat du réseau neuronal, et les résultats de trois algorithmes d'apprentissage : Arbre de décision, Machine à vecteurs de support «SVM » et k plus proche voisin « KNN ». Nous avons choisi ces trois méthodes car elles sont très utilisées dans la littérature, nous avons cité aussi l'environnement de développement,à la fin nous avons présenté notre application web.

Pour que les résultats de cette classification soient réellement comparables, il aurait fallu d'utiliser ces algorithmes sur la même base, et utilisons le même « PC »pour la configuration.

3.2 Les critères d'évaluation

Le critère d'évaluation est un facteur clé à la fois dans l'évaluation de la performance de classification et guidance de la modélisation de classificateur. Pour comparer de façon synthétique les performances des différentes méthodes et de différents outils retenues pour notre étude nous avons calculé : accuracy, la précision, le rappel et score f1.

Les performances de classification des données ont été évaluées par le calcul des vrais positifs (VP), vrais négatifs (VN), faux positifs (FP) et faux négatifs (FN), leurs définitions respectives sont les suivantes :

- VP : diabétique classé diabétique.
- VN : non diabétique classé non diabétique.
- FP : non diabétique classé diabétique.
- FN : diabétique classé non diabétique.

3.2.1 Accuracy

L'accuracy est une mesure de performance pour évaluer les modèles de classifications.
sa formule est la suivante :

$$\text{accuracy} = \frac{VP + VN}{VP + VN + FP + FN} \tag{3.1}$$

3.2.2 La précision (The précision)

La précision est intuitivement la capacité du classificateur à ne pas étiqueter comme positif un échantillon négatif.

La précision est le rapport :

$$\text{précision} = \frac{VP}{VP + FP} \tag{3.2}$$

3.2.3 Le rappel (the recall)

Le rappel est intuitivement la capacité du classificateur à trouver tous les échantillons positifs.

Le rappel est le rapport :

$$\text{recall} = \frac{VP}{VP + FN} \tag{3.3}$$

3.2.4 Score F1

Le score F1 est une mesure globale de l'accuracy d'un modèle qui combine la précision et le rappel par leur moyenne harmonique sa formule est la suivante :

$$\text{Score F1} = \frac{2 * (prcision * rappel)}{prcision + rappel} \tag{3.4}$$

3.3 La configuration du matériel utilisé dans notre expérimentation

— Un Pc portable Samsung i5 CPU 2.60 GHZ.
— RAM de taille 8 GO.
— Système d'exploitation Windows 10, 64 bits.

3.4 Les expérimentations

Dans cette partie nous avons effectué deux expérimentations sur l'ensemble de données utilisées. Dans la première expérimentation nous avons utilisé comme entrée les huit attributs de la base Pima, et pour la deuxième expérimentation on a pris comme entrée seulement quatre attributs de la base (Glucose, insuline, indice de masse corporelle, Diabètes Pedigree Fonction).

3.4.1 Expérimentation 1

Dans notre première expérimentation nous avons utilisé un perceptron multicouche avec la topologie de huit neurones dans la couche d'entrée(les paramètres de la base Pima), quatre neurones dans la couche cachée et la couche de sortie ,avec 200 itérations impliquées pendant le processus de formation du réseau neuronal et une taille de test =0,2.

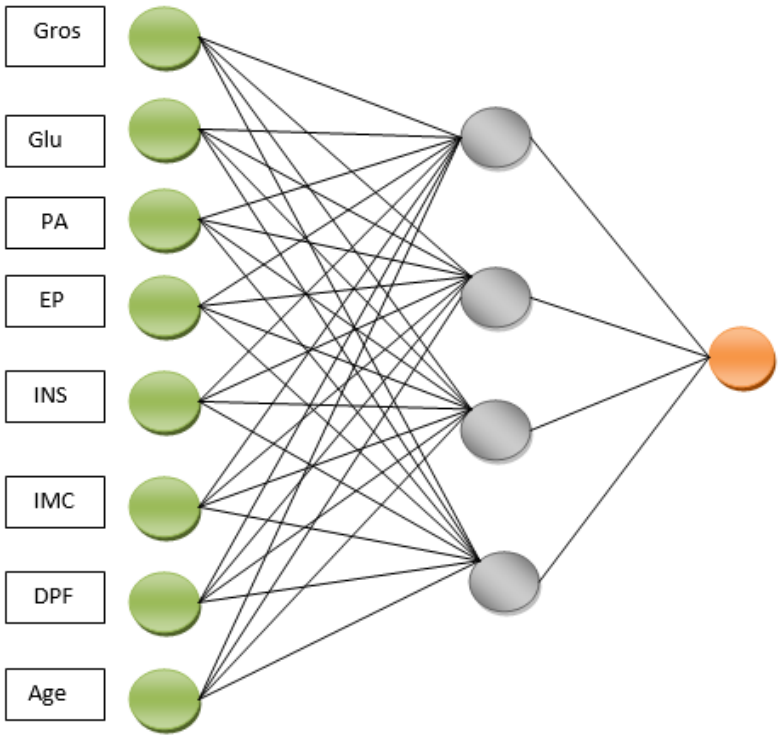

Figure 3.1: Architecture utilisée dans l'expérimentation 1.

Les résultats de cette expérience sont présentés dans le tableau en dessous :

Algorithme / Critéres	**RNA**
Précision	0,679245
Rappel	0,765957
Score F1	0,720000
Accuracy	0,818182

Table 3.1: Résultats obtenus pour l'expérimentation 1

D'après le tableau 3.1 on peut observer que :

— La précision du système est moyenne ce qui veut dire que notre système ne fait pas une bonne reconnaissance pour les données négatives. Donc "33 pourcents " des patients non diabétiques ont été reconnus comme des diabétiques, ce qui ne peut pas générer un risque majeur pour la santé du patient.

— Le rappel est élevé par rapport la précision ce qui veut dire que notre système a fait un bon apprentissage des données positives. . Donc lorsqu'un patient est diabétique notre modèle le détecte avec succès.

— Avec ces performances, nous pouvons dire que le modèle a donné un taux de classification élevée, ”81 pourcents” des cas correctement classés.

3.4.2 Expérimentation 2

Dans le but de comparait la performance de l'expérimentation précédente, nous avons utilisé une architecture de quatre neurones dans la couche d'entrée, quatre neurones dans la couche cachée et un neurone de sortie avec une taille de test =0.2 et un maximum d'itération=200.

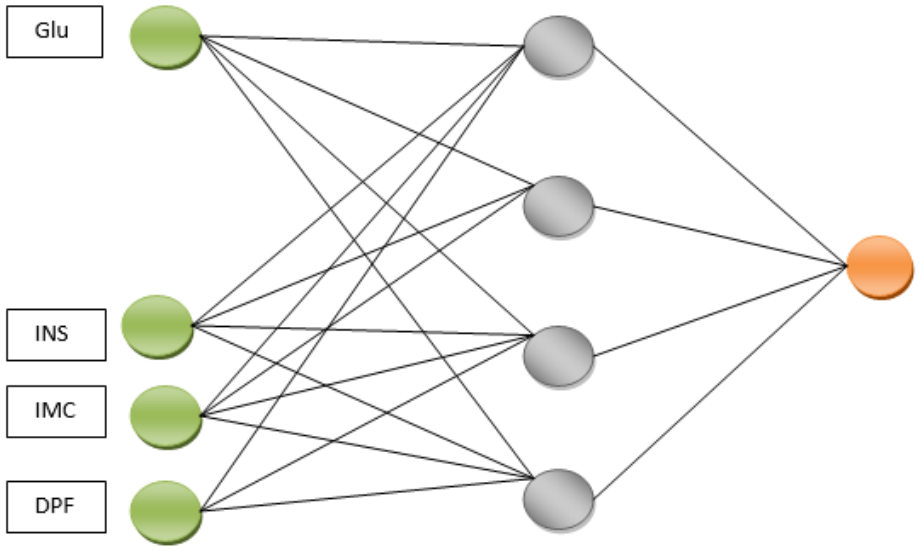

Figure 3.2: Architecture utilisée dans expérimentation 2.

Dans cette expérience on a sélectionné les variables les plus pertinentes (Glu : Glucose, INS : Insuline, IMC : la masse, DPF : L'hérédité). Par ce que le changement de la concentration du glucose et d'insuline sont les paramètres les plus utilisés pour le diagnostic de cette maladie et si nous revenons aux causes du diabète, nous remarquons que la majorité des gens diabétiques souffrent du problème du surpoids et que le facteur génétique est responsable de la plupart des cas des défaillances du pancréas qui est l'anomalie principale d'un développement d'un diabète.

Les résultats de cette expérimentation sont représentés dans le tableau 3.2 :

Critéres \ Algorithme	**RNA**
Précision	0.722222
Rappel	0.553191
Score F1	0.626506
Accuracy	0.798701

Table 3.2: Résultats obtenus pour l'expérimentation 2

— La précision du système est un peu élever ce qui veut dire que notre système a une bonne reconnaissance pour les données négatives. Donc "28 pourcents" des patients non diabétiques ont été reconnus comme des diabétiques, ce qui ne peut pas générer un risque majeur pour la santé du patient.

— Le rappel du système est très faible ce qui veut dire que le système a fait une mauvaise reconnaissance des données positives. Donc beaucoup de patients diabétiques ont été reconnu comme non diabétiques. Ce qui peut générer un risque majeur pour la santé du patient.

— Avec ces performances, nous pouvons dire que le modèle a donné un taux de classification un peu élevé, "79 poucents" des cas correctement classés.

En comparant les résultats des deux expérimentations, nous remarquons que la méthode proposée a diminué le nombre de variables d'entrée de plus de" 50 poucents" et que le nombre de connexions s'est diminué de "50 pourcents", ce qui n'a pas donné des bonnes performances du classifieur, donc ont conclu que la première expérimentation avec les huit paramètres mieux que la deuxième.

3.5 Comparaisons par rapport au nombre d'epochs

Dans cette partie nous avons réalisé une étude comparative entre les résultats des expérimentations obtenus par le changement de l'hyper paramètre :

— **Nombre itération (epochs) :** représente le nombre de fois que l'ensemble de données sera transmis via le réseau de neurones.

Critéres / Paramétres	**Précision**	**recall**	**F1-Score**	**Accuracy**
epochs=50 test size=0.2	0,619048	0 ,553191	0 ,584270	0,759740
epochs=100,testsize=0.2	0,627906	0 ,574468	0 ,600000	0,766233
epochs=200,testsize=0.2	0,679245	**0,765957**	0 ,720000	0,818182
epochs=300,testsize=0.2	0,719048	0 ,595744	0 ,684570	0,753246
epochs=400,testsize=0.2	0,810811	0, 638298	0,714286	0,844156
epochs=500,testsize=0.2	0, 839048	0 ,595744	0 ,738695	0, 849740

Table 3.3: Tableau des résultats de comparaisons par rapport au nombre d'epochs.

Pour choisir le meilleur nombre d'itération (épochs) pour notre programme on a fait plusieurs essais de combinaison de ce hyper paramètre avec la taille de teste. Nous avons collecté tous les résultats obtenus dans ce tableau ; puisque le rappel c'est le paramètre le plus sensible parce que c'est dangereux qu'on donne un résultat négatif à une personne qui a en réalité un résultat positif, on a cherché le plus élevé pour avoir le meilleur couple.

Alors d'après l'analyse des résultats on a remarqué que le rappel (en gras) est élevé lorsque le nombre d'itérations est 200 et la taille de test est 0.2, donc lorsqu'un patient est diabétique ce modèle le détecte avec succès. Cette combinaison donne aussi une accuracy parmi les meilleurs et une bonne précision.

— **Graphe représente l'évolution de la précision avec le nombre d'itérations :**

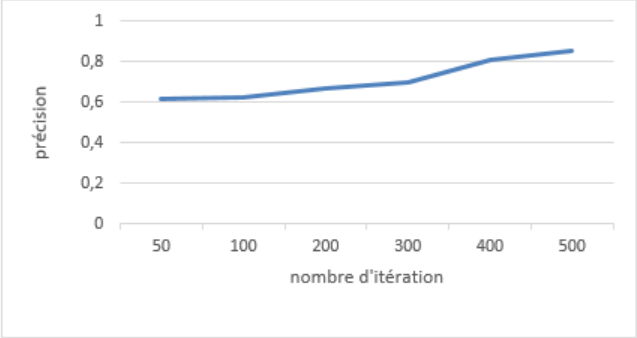

Figure 3.3: Courbe d'évaluation de la précision avec le nombre d'itérations.

D'après la figure 3.3 on remarque que plus le nombre d'itérations augmente, la précision de l'apprentissage de notre modèle sera élevée ; ceci reflète qu'à chaque itération le modèle apprend plus d'informations.

3.6 Comparaisons avec d'autres algorithmes d'apprentissage

Le tableau ci-dessous montre les différents résultats obtenus sur les quatre algorithmes d'apprentissage : Réseau de neurones artificiel (RNA), Arbre de décision (AD), Machine à vecteurs de support (SVM), k plus proches voisin(KNN) :

Algorithmes / Critéres	**RNA**	**AD**	**SVM**	**KNN**
Accuracy	0,818182	0,675324	0,770562	0 ,770562
Précision	0,679245	0,569444	0,847826	0,735294
Reppel	0,765957	0,482352	0,458823	0,588235
Score F1	0,720000	0,522292	0,595419	0,653594

Table 3.4: Tableau de comparaison des résultats des quatre algorithmes.

3.6.1 Discussion des résultats

Dans cette partie on a fait une comparaison de notre algorithme avec d'autres algorithmes d'apprentissage et pour ça on a choisi trois algorithmes : Arbre de décision (AD), Machine à vecteurs de support (SVM), k plus proches voisin(KNN). Et on a choisi aussi pour cette comparaison quatre paramètres : Accuracy qui est le taux de classification, Précision, Rappel, score F1.

a) **Accuracy :**

Pour l'accuracy on a trouvé que SVM et KNN ont la même accuracy qui est "77 pourcents" tandis que pour AD a un taux égal "67 pourcents", donc pour ces trois algorithmes on peut conclus que SVM et KNN sont plus performent qu'Ad. Et pour RNA on a trouvé qu'il a "81 pourcents" comme taux de classification alors c'est le meilleur par rapport aux trois autres algorithmes utilisés dans cette comparaison.

b) **Précision :**

Pour la précision on a "67 pourcents "pour RNA, "56 pourcents" pour AD, "84 pourcents " pour SVM et" 73 pourcents" pour KNN ; par ces résultats on remarque qu'AD a la mauvaise précision par rapport aux autres tandis que le SVM a la précision la plus élevée, et pour notre algorithme RNA on a une précision plus que la moyenne.

c) **Rappel :**

Pour le rappel le SVM a "45 pourcents" qui sont le résultat le plus mauvais, il suit par AD qui a "48 pourcents". Donc si on basant sur le rappel on peut dire que le SVM et AD sont des mauvais algorithmes pour la classification binaire car le rappel est moins "50 pourcents", il est très faible. On a trouvé aussi le rappel de KNN et moyenne, il égale "58 pourcents", et le rappel le plus élevé c'est le rappel de RNA qui est égale" 76 pourcents".

d) **Score F1 :**

Pour le score F1 on a le moins score obtenu par l'algorithme AD qui est égale "56 pourcents", et "59 pourcents" pour le SVM, le KNN a un score un peu plus élevé par rapport à AD et

SVM égale ”65 pourcent” tandis que le meilleur score est ”72 pourcenst” qui est obtenu par RNA. Alors avec ces résultats et cette comparaison entre eux on a conclu que l'algorithme AD est le mauvais algorithme pour la classification binaire, on a trouvé aussi que le SVM a la précision la plus élevée mais elle a un rappel le plus mauvais et moins que la moyenne par rapport aux autres algorithmes.

Tandis que l'algorithme RNA c'est le meilleur par rapport aux autres algorithmes grâce à l'accuracy qui est la plus élevée ce qui rendre RNA le plus performent que les autres. Il a aussi le meilleur rappel qui est un paramètre très sensible parce que c'est dangereux qu'on dise à une personne diabétique que son résultat est négatif. L'algorithme a une précision plus que la moyenne et un meilleur score.

3.7 Conception et implémentation

3.7.1 L'environnement de développement

Après avoir achevé notre conception nous avons donné les environnements pour la réalisation de notre travail :

3.7.1.1 Anaconda

Anaconda est une distribution et multiplate forme libre et open source des langages de programmation appliqués au développement d'applications dédiées à la science des données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul scientifique). Il a plusieurs analogies avec pip et virtualenv, mais il est conçu pour être plus indépendant de python.[13]

Figure 3.4: Logo de l'environnement ANACONDA.

3.7.1.2 Spyder

Spyder (Scientific PYthon Development Environment), c'est un environnement de développement scientifique Python inclus dans Anaconda. Il est intégré pour le langage Python avec des fonctionnalités avancées d'édition, de test interactif, de débogage et d'introspection. Il prise en charge plusieurs bibliothèques comme IPython pour le calcul numérique, NumPy pour les fonctions de l'algèbre linéaire, SciPy pour le traitement des signaux ou d'images et matplotlib,. . . etc.[14]

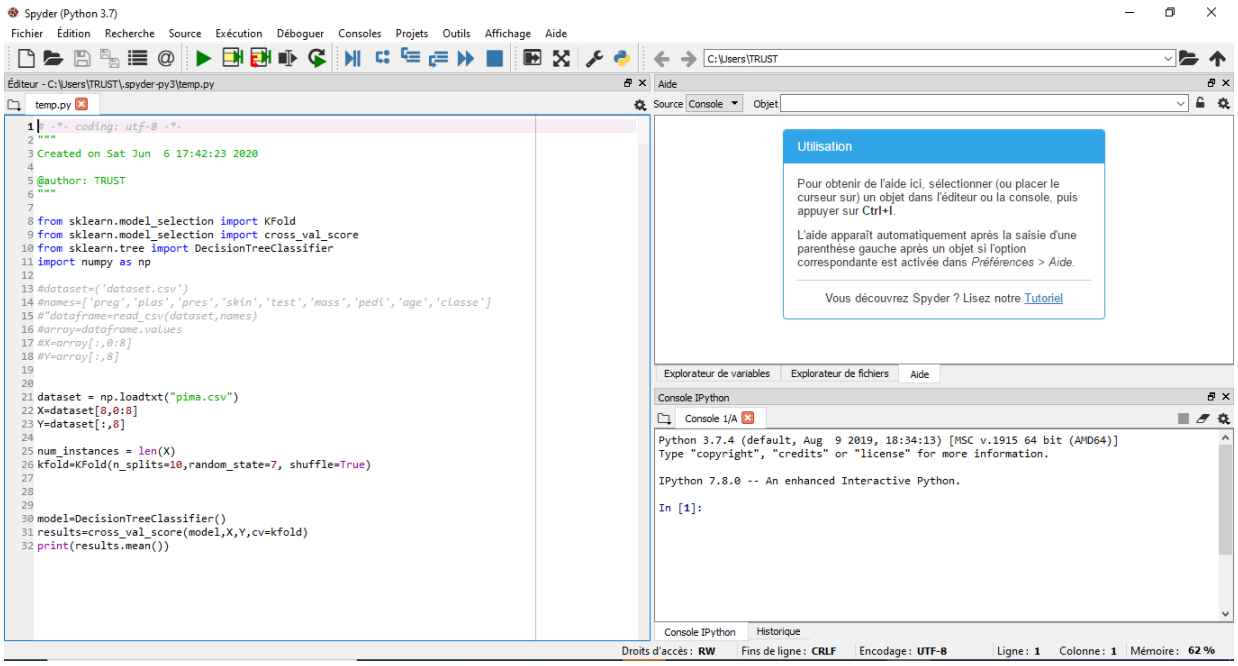

Figure 3.5: L'environnement spyder.

3.7.1.3 Flask

Flask est un Framework qui utilise pour faire la compatibilité entre différents types d'application et entre différents serveurs Web. Il a l'avantage d'être simple à apprendre tout en pouvant profiter de la puissance de python permette de faciliter la création des sites web interactifs. Flask s'inspire son mode de fonctionnement du fonctionnement de HTTP et permet d'écrire un programme qui remplace les requêtes « GET http » faites par le navigateur.[37]

Figure 3.6: Logo de l'environnement Flask.

3.7.2 Structure du projet

Pour réaliser ce projet nous avons divisé notre travail à quatre parties :

1) **model.py :** Il contient du code pour le modèle d'apprentissage automatique afin de prédire si le patient est diabétique ou non diabétique.

2) **app.py :**Il contient des API Flask qui reçoivent les détails des patients via des appels GUI ou API, calcule la valeur prédite en fonction de notre modèle et la renvoie.
3) **New-classification.py :** Cela utilisé pour effectuer la classification de notre application.
4) **HTML / CSS :** Il contient le modèle HTML et le style CSS pour permettre à l'utilisateur d'entrer les détails et affiche la probabilité d'avoir un diabète ou non.

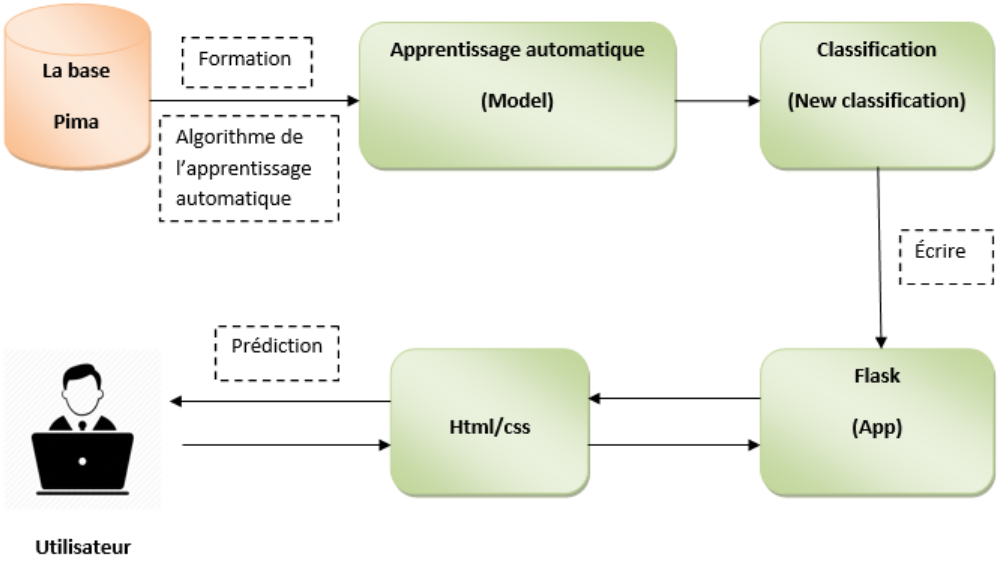

Figure 3.7: Structure de notre projet.

3.7.3 Présentation de l'application web

Après la création de notre architecture de réseau de neurones et la sélection des hyperparamètres par les comparaisons qu'on a faite pour avoir un modèle performant ; nous avons implémenté une IHM pour que les visiteurs de notre site web peuvent communiquer et interagissent avec notre machine automatique. Pour que notre site web approximé au secteur médical on a utilisé trois couleurs principales :

1) **Blanc :**
Le blanc est la couleur par excellence de la pureté et de l'apaisement. C'est une couleur qui sied parfaitement à l'univers médical, que ce soit dans le monde pharmaceutique, technologique ou même cosmétique. Elle évoque parfois le luxe avec la simplicité. Il est difficile de s'écarter d'une utilisation au moins partielle du blanc dans un site médical.
2) **Bleu :**
Le bleu est une couleur fraîche évoque l'assurance, la bienveillance et l'expertise. C'est la couleur phare des grands groupes et des entreprises pharmaceutiques. Elle témoigne d'un caractère empathique et alimente la notoriété d'une entreprise compétente et à l'écoute de ses patients.

3) **Jaune :**
Associé à l'énergie, il fonctionne particulièrement bien dans la partie publicitaire de l'environnement de santé. Elle est parfaitement convenue à l'identité visuelle d'un produit cosmétique ou d'un établissement médical.

3.7.3.1 Fenêtre principale

Cette fenêtre s'ouvre après le lancement du serveur, elle permet au client d'accéder à des pages qu'on a proposé pour qu'il ait des informations sur la maladie diabète ou sur des quelques médicaments ou bien des aliments qui peuvent l'aide ; et un bouton pour aller à la page de diagnostic.

On a fait aussi deux autres pages à-propos pour dire aux visiteurs qui somme nous et contact pour que le client nous contacter s'il a une question ou trouve un problème.

Figure 3.8: Fenêtre principale.

3.7.3.2 Première fenêtre

Dans cette page on a mis quelques informations sur le diabète pour que le visiteur avoir une idée sur cette maladie chronique.

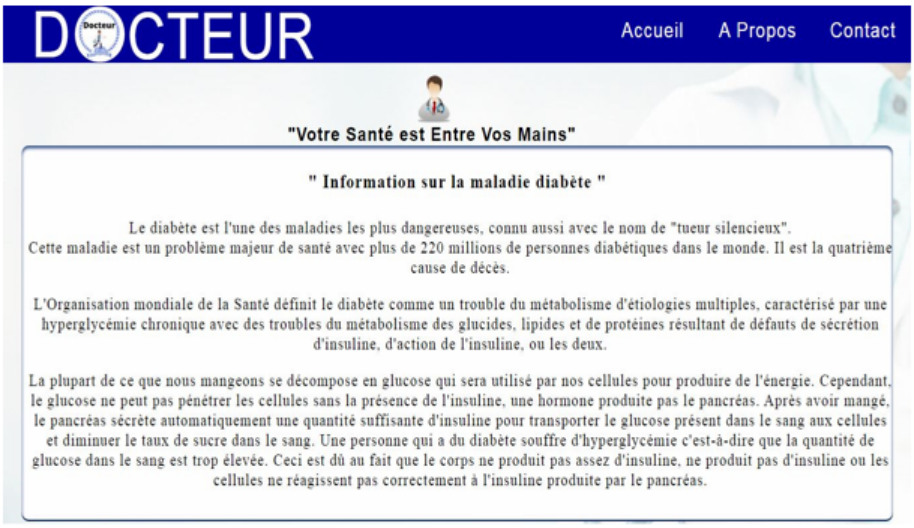

Figure 3.9: Première fenêtre.

3.7.3.3 Deuxième fenêtre

Cette page contient quelques exemples des médicaments qui sont progrès aux malades diabétique.

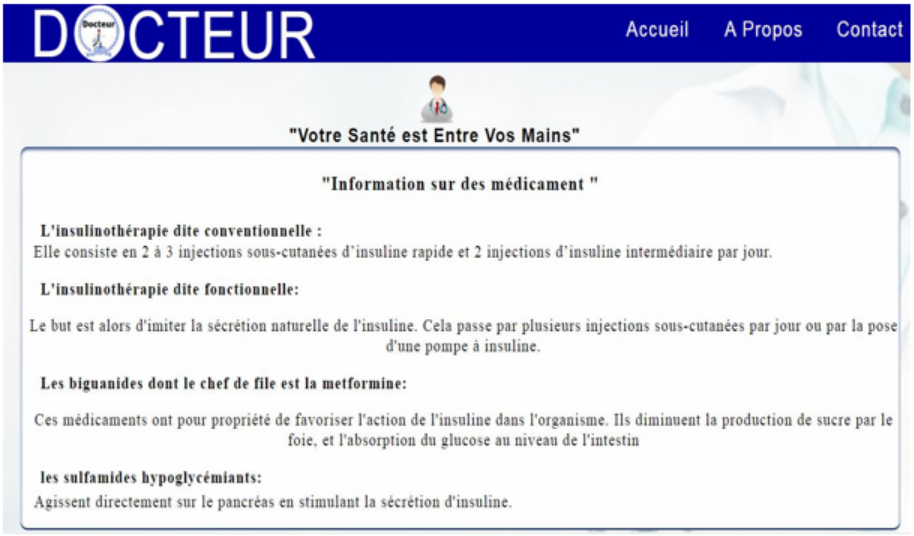

Figure 3.10: Deuxième fenêtre.

3.7.3.4 Troisième fenêtre

Dans cette fenêtre on propose trois alimentaires nécessaires pour soutenir la santé des malades diabétiques ou pour aider les gens de n'avoir pas cette malade dangereuse.

Figure 3.11: Troisième fenêtre .

3.7.3.5 La fenêtre de diagnostic

Cette fenêtre apparaît par le clic sur le bouton "Diagnostique du diabète" dans la page principale ; elle permet de lancer le système de diagnostic après que le visiteur remplit les huit champs par leurs informations ,puis faire un clic sur le bouton "Diagnostique" et le résultat sera apparu dans le fond du cadre en bleu.

Figure 3.12: La fenêtre de diagnostic .

3.8 Conclusion

La conclusion qu'on peut tirer de cette étude comparative est que l'architecture qui est composé de tous les attributs de la base est meilleure par rapport à l'architecture où on a diminué le nombre des variables d'entrée car le risque de ne pas reconnu les cas positifs et élevés.Ensuite on a choisi les trois algorithmes SVM, KNN, AD pour les comparer avec notre système, on a trouvé que le système basé sur SVM a fait une reconnaissance des cas négatifs plus mieux que notre système ,mais il donne un pourcentage le plus faible pour la reconnaissance des cas positifs qui est le plus nécessaire etde diminué sa performance.

Alors que notre RNA a le taux de classification le plus élevé que les autres, il a aussi un rappel majeur ce qui nous aide à éviter le risque qu'un cas positif ait un résultat négatif. Donc on a conclu que par rapport à tous les systèmes qu'on a choisis pour cette étude comparative notre système est le plus performant.

Conclusion générale

Ces dernières années, avec le développement des systèmes informatiques et logiciels complexes, l'intelligence artificielle s'est immiscée dans de nombreux domaines et elle a pénétré dans presque tous les secteurs pour réaliser des systèmes experts aide à la décision basée sur l'apprentissage automatique et les réseaux de neurones. Bien que la décision d'un expert soit un facteur important dans le diagnostic médical, mais les systèmes experts et les différentes techniques d'intelligence artificielle ont prouvé leur efficacité d'aider les experts dans le domaine médical. Ce qui nous avons stimulé à modéliser cette pratique.

Dans ce travail de mémoire tout d'abord nous avons commencé dans le premier chapitre par la présentation de l'intelligence artificielle et la machine automatique et on a discuté sur des notions fondamentales de plusieurs algorithmes utilisés pour l'apprentissage automatique.

Ensuite dans le chapitre suivant on a décrit le réseau de neurones artificiel qu'on a choisi pour résoudre notre problématique d'apprentissage de la machine automatique et nous avons indiqué la base de données médicale utilisée par notre apprentissage qui est « Indians Diabètes Pima ».

Dans le troisième et le dernier chapitre nous avons fait des expérimentations et on a présenté et discuté leurs résultats. En premier lieu on a fait une comparaison entre deux architectures par le changement de nombre des attributs d'entrer dans le réseau ; en deuxième lieu on a choisi trois algorithmes d'apprentissage les plus utilisés qui sont AD (Arbre de décision), KNN (K plus proche voisin), et SVM (Machine à vecteurs de support) et nous avons comparé leurs résultats avec le résultat de notre réseau neuronal ; et on a terminé par la présentation d'IHM de notre système.

Globalement, cette étude a permis d'exposer une solution pour la classification de données dans le domaine médical, pouvant en cela contribuer et aider plusieurs experts dans ce domaine pour le traitement, et aussi aide les gens à faire un diagnostic médical à distance sans aller chez un docteur.

Comme perspectives, nos objectifs se focalisent sur l'enrichissement de notre site web avec d'autres domaines médicales comme la médecine générale , le covid19 la Maladie du siècle ...

Annexe1

Apprentissage des réseaux neurone Artificiel :

— **Rétro propagation de l'erreur :**
Mécanisme par lequel les erreurs d'interprétation, calculées à la sortie d'une ou de plusieurs couches de neurones d'un réseau de neurones artificiels, produisent des signaux qui sont transmis vers les neurones qui ont contribué précédemment à créer des écarts, afin que des correctifs soient apportés en ajustant les coefficients synaptiques ou les biais responsables.[38]

L'algorithme qu'on désigne souvent par rétropropagation est le plus populaire parmi les techniques d'apprentissage des réseaux multicouches. L'algorithme de rétropropagation est basé sur la généralisation de la règle de Pinède [16] en utilisant une fonction d'activation sigmoïde. Le réseau utilisé est un réseau à couches où chaque neurone est connecté à l'ensemble des neurones de la couche suivante. Le principe de cet algorithme est la propagation d'un signal provenant des nœuds d'entrées vers la sortie et ensuite on propage l'erreur commise de la sortie vers les couches internes jusqu'à l'entrée afin de calculer la nouvelle matrice des poids.[17]

— **Algorithmes d'optimisation (gradient) :** L'algorithme de minimisation utilisé est celui de la descente du gradient stochastique [18]. Le gradient d'une fonction en un point est défini comme le vecteur qui pointe vers le maximum local de cette fonction le long de la pente la plus abrupte. Une technique de minimisation de l'erreur quadratique instantanée selon le négatif du gradient assure donc convergence relativement rapide vers une erreur minimum (à tout le moins localement).

L'algorithme de descente de gradient stochastique consiste donc à exprimer le gradient en fonction des poids de connexion du réseau et à trouver l'amplitude et le sens des changements de poids qui minimisent le gradient de la fonction d'erreur instantanée pour la forme Xk présentée à l'entrée du réseau.

a) **Avantages :**
- Taux d'erreur généralement bon.
- Outil disponible dans les environnements de datamining .
- Classification rapide (réseau étant construit) .
- Combinaison avec d'autres méthodes (ex : arbre de décision pour sélection d'attributs).

b) **Inconvénients :**
- Apprentissage très long.
- Plusieurs paramètres (architecture, coefficients synaptiques s, ...) .
- Pas facile d'incorporer les connaissances du domaine .
- Traitent facilement les attributs numériques et binaires .
- Évolutivité dans le temps (phase d'apprentissage).

Annexe2

Définition

— **Python :** Python est un langage de programmation open source (gratuit) de haut niveau développé en 1989 par Guido van Possum [19] et publié pour la première fois en 1991. C'est un langage multiplate-forme fonctionné sur nombreux systèmes d'exploitation. Il a une syntaxe simple, lisible, et très puissante en termes de production, caractérisé par un typage dynamique où le typage fait automatiquement lors de l'exécution du programme [20] ce qui permet une grande flexibilité et rapidité de programmation.

— **Tensorflow :** Tensorflow est une plateforme open source de bout en bout pour l'apprentissage automatique développé par l'équipe Google Brian Elle implémente des méthodes basées sur le principe des réseaux de neurones profonds. Il dispose d'un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires qui permet aux chercheurs de faire évoluer l'état de l'art en machine learning et aux développeurs de créer et de déployer facilement des applications propulsées par machine Learning. [21] [22]

— **Keras :** Est une librairie Python open source gratuite puissante et facile à utiliser pour développer et évaluer des modèles d'apprentissage en profondeur . Il encapsule l'accès aux fonctions proposées par plusieurs librairies de machine Learning, et enveloppe les bibliothèques de calcul numérique efficaces comme Theano et Tensorflow. Il a été développé dans le but de permettre une expérimentation rapide. Pouvoir passer de l'idée au résultat avec le moins de retard possible est la clé d'une bonne recherche. [23]

L'avantage de ceci est principalement que vous pouvez démarrer avec les réseaux de neurones de manière simple et amusante.

— **Theano**
Theano est un projet open source développé principalement par l'Institut des algorithmes d'apprentissage de Montréal (MILA) de l'Université de Montréal .

Theano est une bibliothèque Python. Il vous permet de définir, d'optimiser et d'évaluer expressions mathématiques, en particulier celles qui sont utilisées dans le développement du modèle d'apprentissage automatique. Theano lui-même ne contient aucun modèle ML prédéfini ; cela facilite simplement son développement. [24]

• Il est particulièrement utile pour les tableaux multidimensionnels.
• Il s'intègre parfaitement à NumPy, qui est un package fondamental et largement utilisé pour calculs scientifiques en Python.
• Theano facilite la définition d'expressions mathématiques utilisée dans le développement ML. Tel (les expressions impliquent généralement l'arithmétique matricielle, la différenciation, le calcul de gradient, etc.)

— **NumPy :** NumPy est le package fondamental pour le calcul scientifique avec Python. Il s'agit d'une bibliothèque logicielle libre et open source utilise pour la création des tableaux multidimensionnels avec des opérations rapides sur les tableaux, y compris mathématique, logique, manipulation de forme, tri, algèbre linéaire de base, opérations statistiques de base, simulation aléatoire...etc .[25]
Il permet de créer directement un tableau depuis un fichier ou sauvegarder un tableau dans un fichier, et effectuer des manipulations sur les vecteurs, matrices et polynômes.

— **Scikit-learn :** Scikit-learn est une bibliothèque libre de Python introduit pour à l'apprentissage automatique supervisés et non supervisés. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de classification, et les machines à vecteurs de support (SVM).[26]

— **Pandas :** Pandas est une bibliothèque libre de Python son nom dérive des données de panel, un terme courant pour les ensembles de données multidimensionnelles rencontrés dans les statistiques et l'économétrie.

  Pandas est un package fournissant des structures de données rapides, flexibles et conçues pour faciliter la manipulation et l'analyse des données et les calculs scientifiques.[27]

Bibliographie

Bibliographie

[1] EZRATTY, Olivier. *Les usages de l'intelligence artificielle*. Olivier Ezratty, 2018.

[2] ATIF, Jamal. Intelligence Artificielle. Intelligence, 2017, no 1/44.

[3] KLÄS, Michael et VOLLMER, Anna Maria. *Uncertainty in machine learning applications : A practice-driven classification of uncertainty. In : International Conference on Computer Safety, Reliability, and Security*. Springer, Cham, 2018. p. 431-438.

[4] SMOLYAKOV, Vadim. *Ensemble learning to improve machine learning results*. Stats et Bots, 2017.

[5] ROKACH, Lior et MAIMON, Oded. *Top-down induction of decision trees classifiers-a survey*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2005, vol. 35, no 4, p. 476-487.

[6] CAPPONI, Cécile. *Arbres de décision. M2 mass, Université Aix-Marseille*. Cité, p. 34.

[7] Belghaba , Brahim ; Boukhris, Mohamed,*l'apprentissage profond(deep learning) pour la classification et la recherche d'image par le contenu*,2017.

[8] SORIN, Fabrice, BROUSSARD, Lionel, et ROBLIN, Pierre. Régulation d'un processus industriel par réseaux de neurones. Techniques de l'ingénieur. Informatique industrielle, 2001, vol. 2, no S7582, p. S7582. 1-S7582. 13.

[9] HOUMADI, Benamar. Étude exploratoire d'outils pour le Data Mining. 2007. Thèse de doctorat. Université du Québec à Trois-Rivières.

[10] MERZOUKA, Nouressadat. Etude des performances des réseaux de neurones dynamiques à représenter des systèmes réel : une approche dans l'espace d'état. 2018. Thèse de doctorat.

[11] SRIVASTAVA, Nitish, HINTON, Geoffrey, KRIZHEVSKY, Alex, et al. Dropout : a simple way to prevent neural networks from overfitting. The journal of machine learning research, 2014, vol. 15, no 1, p. 1929-1958.

[12] BONAZZA, Pierre. *Système de sécurité biométrique multimodal par imagerie, dédié au contrôle d'accès*. 2019. Thèse de doctorat. Bourgogne Franche-Comté.

[13] LANDRUM, Greg. *Rdkit documentation*. Release, 2013, vol. 1, p. 1-79.

[14] RAYBAUT, Pierre. *Spyder-Documentation*. Available online at : pythonhosted. org, 2009.

[15] MARTY, Jean-Marc, WENZEK, Guillaume, SCHMITT, Eglantine, et al. *Analyse d'opinions de tweets par réseaux de neurones convolutionnels*. Actes de DEFT, Caen, France : TALN, 2015.

[16] PINEDA, Fernando J. *Generalization of back-propagation to recurrent neural networks. Physical review letters*, 1987, vol. 59, no 19, p. 2229.

[17] RUMELHART, David E., HINTON, Geoffrey E., et WILLIAMS, Ronald J. *Learning internal representations by error propagation*. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[18] BLAYO, François et VERLEYSEN, Michel. Les réseaux de neurones artificiels. 1996.

[19] P. Poulain et P. Fuchs, *Cours de Python*, 2019.

[20] Y. Derfoufi , *Formation en langage Python*, 2019.

[21] SERGEEV, Alexander et DEL BALSO, Mike. *Horovod : fast and easy distributed deep learning in TensorFlow*. arXiv preprint arXiv :1802.05799, 2018.

[22] ABADI, Martín, BARHAM, Paul, CHEN, Jianmin, et al. *Tensorflow : A system for large-scale machine learning. In : 12th USENIX symposium on operating systems design and implementation (OSDI 16)*. 2016. p. 265-283.

[23] Claude. Keras documentation, 2015.

[24] 19. James, BREULEUX, Olivier, BASTIEN, Frédéric, et al. Theano : A CPU and GPU math compiler in Python. In : Proc. 9th Python in Science Conf. 2010. p. 3-10.

[25] OLIPHANT, Travis E. A guide to NumPy. USA : Trelgol Publishing, 2006.

[26] 21. PEDREGOSA, Fabian, VAROQUAUX, Gaël, GRAMFORT, Alexandre, et al. Scikit-learn : Machine learning in Python. the Journal of machine Learning research, 2011, vol. 12, p. 2825-2830.

[27] MCKINNEY, Wes, et al. pandas : a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing, 2011, vol. 14, no 9.

Webographie

[28] Cleuet, valentin, «l'intelligence artificielle et ses axes de rechereches» 2 jullet 2019 [en ligne].available : https ://www.cloud-temple.com/intelligence-artificielle-axes-recherches. [accés le 20 mars 2020].

[29] Kush, «different perspectives on linear regression,» 2018. [En ligne]. [Accès le 17 Mars 2020].

[30] David et varadi,«combining acceleration and volatilty into a Non Linear Filter(NLV),» 3 December 2014. [En ligne]. [Accès le 17 Mars 2020].

[31] Pegliasco,Gaël ,« IInitiation au Machine Learning avec Python - La théorie,» 31 janvier2019 [en ligne].available : https ://makina-corpus.com/blog/metier/2017/initiation-au-machine-learning-avec-python. [accés le 20 mars 2020].

[32] Sanjay,M,«to words data science»,26 Octobre 2018.[En ligne]. Available :https ://towardsdatascience.com/.[Accès le 29 Juillet 2020].

[33] «DeepAI,» [En ligne]. Available : https ://deepai.org/machine-learning-glossary-terms/machine-learning. [Accès le 17 Mars 2020].

[34] «WIkivérsité,» 14 janvier 2019. [En ligne]. Available : https ://fr.wikiversity.org/. [Accès le 21 juillet 2020].

[35] Verseils et Gabriela,«MBA MC LET'S GO DIGITAL,»3septembre 2018[en ligne] .available :http//mbamci.com/machine-learning-et-personnalisation.[accésle 17 mars 2020].

[36] Y. Benzaki, «Mr.Mint Machine Learning made easy,» 12 octobre 2017. [En ligne]. Available : https ://mrmint.fr/data-preprocessing-feature-scaling-python. [Accès le 18 juillet 2020]

[37] Bonaventure et Olivier, «Le framework flask,» 2020. [En ligne]. Available : https ://sites.uclouvain.be/P2SINF/flask.html. [Accès le 18 07 2020].

[38] «Une intelligence artificielle bien réelle : les termes de l'IA,» [En ligne]. [Accès le 12 fevrier 2020].

Résumé

Résumé

L'utilisation de l'intelligence artificielle et de d'analyse de données dans le domaine médical est de plus en plus fréquente afin de minimiser le taux d'erreur qui peut causer le décès du patient, améliorer la qualité et le temps du diagnostic, par l'utilisation des techniques dites intelligentes pour l'aide au diagnostic médical.

Ce projet s'intéresse particulièrement à implémenter une application web d'aide au diagnostic médical à distance du diabète du type 2. Nous avons proposé une architecture de réseau de neurones artificiels adaptée à la base de données de diabète Pima. Nous avons comparé les résultats avec d'autres méthodes de classification, les résultats était satisfaisantes en termes de précision et en temps de classification.

Mots clés : Aide à la décision , Intelligence artificielle, Web service, Classification, Réseau de neurones, Apprentissage automatique, Diabète de type 2, Pima.

Abstract

The use of artificial intelligence and data analysis in the medical field is more and more frequent in order to minimize the error rate that can cause the death of the patient, ameliorate the quality and time of diagnosis, by using so-called intelligent techniques to help with medical diagnosis.
This project is particularly interested in implementing a web application of medical diagnosis of type 2 diabetes in distance. We have suggested artificial neural network architecture adapted to the Pima diabetes database. We compared the results with other classification methods ; the results were satisfactory in terms of precision and classification time.

Keywords : Decision help, Artificial intelligence, Web service, Classification, Neural network, Machine learning, Type 2 diabetes, Pima.

ملخص :

يتزايد استخدام الذكاء الاصطناعي وتحليل البيانات في المجال الطبي من أجل تقليل نسبة الخطأ الذي يمكن أن يسبب وفاة المريض، وتحسين جودة و وقت التشخيص، باستخدام ما يسمى التقنيات المساعدة الذكية مع التشخيص الطبي.

يهتم هذا المشروع بشكل خاص بتنفيذ تطبيق على الأنترنت للتشخيص الطبي لمرض السكري من النوع 2عن بعد. لقد اقترحنا بنية شبكة عصبية اصطناعية تتكيف مع قاعدة بيانات Pima **للسكري. قارنا النتائج مع مناهج التصنيف الأخرى، وكانت النتائج مرضية من حيث الدقة و وقت التصنيف.**

الكلمات الدالة:
المساعدة على اتخاذ القرار، الذكاء الاصطناعي، خدمة الانترنت، التصنيف، الشبكة العصبية، التدريب الأوتوماتيكي، السكري من النوع 2، Pima .