

Prédire l'apparition de maladies chroniques chez des individus à des fins de prévention

Les maladies chroniques sont de plus en plus présentes dans le monde, 71% des décès en 2017 leur sont attribués contre 63% en 2008. Pour cette raison, aider à leur prévention peut avoir un fort impact positif sur la santé mondiale.

Le sujet concerne spécifiquement la prévention des risques de santé chez l'individu. En effet la prédiction de l'apparition d'une maladie chronique va permettre de réduire ce risque en l'incitant par exemple à faire évoluer son hygiène de vie ou à prendre des traitements préventifs.

Positionnement thématique (ETAPE 1)

INFORMATIQUE (Informatique pratique), INFORMATIQUE (Informatique Théorique), MATHEMATIQUES (Mathématiques Appliquées).

Mots-clés (ETAPE 1)

Mots-Clés (en français)	Mots-Clés (en anglais)
<i>Réseaux de neurones</i>	<i>Neural networks</i>
<i>Apprentissage machine</i>	<i>Machine Learning</i>
<i>Régression logistique</i>	<i>Logistic Regression</i>
<i>Descente du gradient</i>	<i>Gradient descent</i>
<i>Retro-propagation</i>	<i>Backpropagation</i>

Bibliographie commentée

Comme l'indique l'Organisation Mondiale de la Santé (OMS) les maladies non transmissibles, aussi appelées chroniques sont un des grands défis des années à venir. Elles représentaient, avant la pandémie actuelle, 71% des décès dans le monde soit 41 millions de morts par an contre « seulement » 63% en 2008 [1] représentant donc une hausse de 12,5% des cas de décès. Loin de toucher seulement les personnes âgées, elles touchaient également les 30/69 ans avec 15 millions de décès par an. Quand il s'agit de prévenir des maladies il est donc particulièrement utile de s'intéresser à celles-ci. En effet en prévenant ne serait-ce que 10% des apparitions de maladies chroniques on pourrait réduire de plus de 7% les décès annuels et avoir 1,5 millions moins de morts chez les moins de 69 ans par an !

On peut alors se demander s'il est possible de les prévoir de manière simple et relativement automatisée afin de rapidement mettre en œuvre les moyens de prévention de manière ciblée sur les individus en ayant particulièrement besoin. On sait aujourd'hui par exemple détecter un cancer de la peau grâce à l'apprentissage machine avec une efficacité équivalente à celle d'un dermatologue [2]. Cependant il s'agit ici de détecter une maladie déjà présente et non pas sa future apparition. C'est bien sûr intéressant, puisque cela permet la mise en place de traitements, mais trop tard pour parler de prévention. On peut aller plus loin et, dans certains cas, prédire l'apparition de la maladie, par exemple : on peut détecter dans le sang la présence de molécules bien particulières

plus de dix ans avant l'apparition du diabète de type I [3]. Il s'agit cependant ici de faire une analyse de sang très spécifique pour détecter ces molécules peu communes. Il est difficilement envisageable de généraliser ce genre de test à l'ensemble de la population.

Il peut être intéressant d'utiliser une approche faisant appel à des critères physiologiques assez commun dans les tests sanguins comme le taux de glucose ou d'insuline et de l'associer à des critères simples comme le nombre de malades dans la famille proche de la personne et d'utiliser ces informations pour prédire l'apparition du diabète. L'utilisation de telles données issues de tests sur la population féminine indienne-américaine des États-Unis d'Amérique a donné des résultats satisfaisants grâce à l'entraînement d'un algorithme d'apprentissage machine appelé ADAP avec 76% d'exactitude sur la prédiction de l'apparition du diabète de type II dans les cinq ans qui suivent [4]. On pourrait étudier si cet algorithme peut donner de meilleurs résultats à partir de données contenant plus d'échantillons que dans l'entraînement initial ou encore si l'on peut obtenir des résultats similaires avec une autre maladie chronique comme le diabète de type I. Cependant les auteurs de cet article suggèrent que d'autres algorithmes d'apprentissage machine tels que la régression logistique ou les réseaux de neurones devraient être essayés pour comparer leurs résultats à partir des mêmes données. D'autre part, les progrès en matière d'apprentissage machine depuis cette étude initiale notamment en termes de fonction d'activation des réseaux neuronaux [5] et les performances accrues des ordinateurs leur permettant d'entraîner des réseaux plus profonds semblent indiquer que l'implémentation d'un réseau de neurones possédant un nombre de couches suffisante et avec les bonnes fonctions d'activation puisse donner des résultats supérieurs aux 76% de l'étude initiale.

Problématique retenue

Dans quelle mesure peut-on prédire à partir d'un certain nombre d'informations physiologiques et historiques sur une personne et d'un algorithme d'apprentissage machine qu'elle sera amenée à déclarer une maladie chronique telle que le diabète avec une exactitude supérieure à 80% ou plus.

Objectifs du TIPE

Je compte repartir des données utilisées dans l'article [4] qui obtient une exactitude de prédiction de l'apparition du diabète à 5 ans pour la population donnée de 76%. Je compte ensuite étudier ces données pour mieux les comprendre puis implémenter en Python des algorithmes d'apprentissage machine tels que la régression logistique ou les réseaux de neurones. Et d'ensuite entraîner ces algorithmes avec différents paramètres et fonctions d'activation afin d'atteindre et de si possible dépasser une exactitude de prédiction de 80%.

Références bibliographiques (ETAPE 1)

[1] OMS : Rapport sur la situation mondiale des maladies non transmissibles :

https://www.who.int/nmh/publications/ncd_report-summary_fr.pdf, 2010

[2] GAUTIER CARIOU : L'apprentissage profond bouleverse les sciences : *La Recherche, mensuel* 529, novembre 2017

[3] ABNER LOUIS NOTKINS : Prédire les maladies auto-immunes : *Pour la science*, n°355, mai 2007

- [4] JACK W. SMITH, BST, JE EVERHART, MD, MPH , WC DICKSONT, WC KNOWLER, MD, DRPH , RS JOHANNES, MD,MS : Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus : *Proceedings of the Annual Symposium on Computer Application in Medical Care, 1988*
- [5] FRÉDÉRIC SUR : Introduction à l'apprentissage automatique, Chapitre 9 : Les réseaux de neurones artificiels : *Tronc commun scientifique FICM 2A École des Mines de Nancy, 2021-2022*, https://members.loria.fr/FSur/enseignement/apprauto/poly_apprauto_FSur.pdf