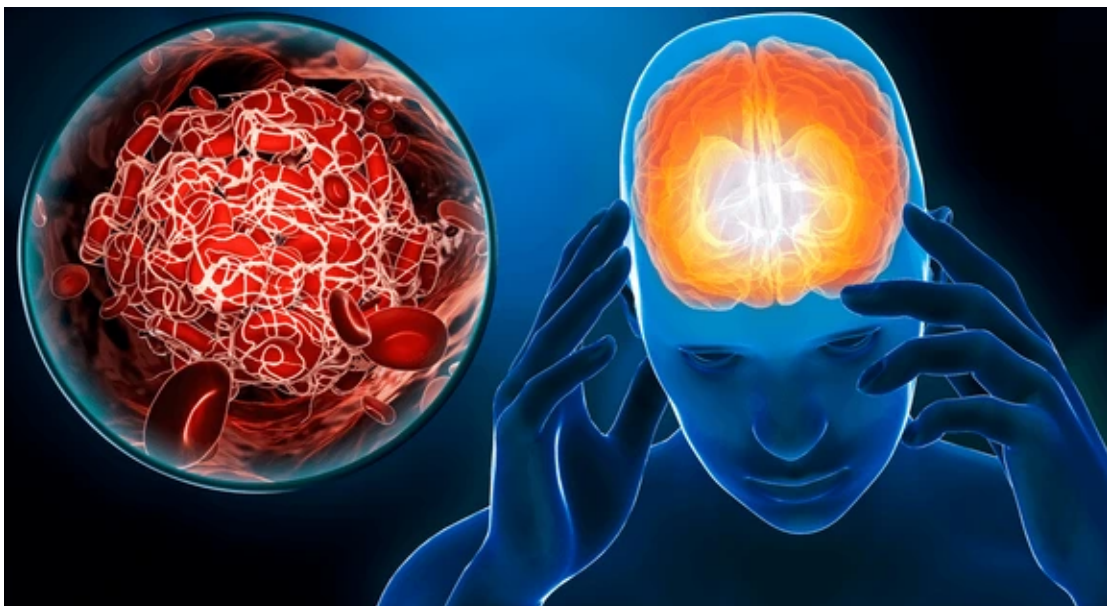


Stroke Alert

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.

Perte soudaine des fonctions du cerveau



Un accident vasculaire cérébral ou AVC, communément appelé « attaque cérébrale », est une perte soudaine d'une ou plusieurs fonctions du cerveau. Il est provoqué :

- Par un arrêt brutal de la circulation sanguine à l'intérieur du cerveau. C'est l'accident vasculaire cérébral ischémique. L'arrêt de la circulation du sang dans une artère ne permet plus un apport suffisant en oxygène et en éléments nutritifs. Cela entraîne la mort des cellules cérébrales, au niveau de la zone du cerveau touchée : c'est l'infarctus cérébral ;



- Ou par la survenue d'une hémorragie intracérébrale. C'est l'AVC hémorragique.

La gravité de l'accident vasculaire cérébral dépend de la localisation et de l'étendue des zones cérébrales touchées.

Chiffres clés concernant les accidents vasculaires cérébraux en France :

En 2021, 121 940 personnes ont présenté un AVC aigu (52,6 % étaient des hommes et 47,4 % des femmes).

Parmi les patients pris en charge pour un AVC aigu, certains présentaient une ou plusieurs autres maladies cardiaques ou facteurs du risque cardiovasculaire :

- **34 % ont des troubles du rythme cardiaque ;**
- **23 % ont une forme de diabète ;**
- **17 % une maladie coronaire chronique ;**
- **8% une artériopathie des membres inférieurs ;**
- **8% une maladie des valves cardiaques.**

En 2021, 849 300 personnes ont été suivies pour des séquelles d'AVC (51,8 % étaient des hommes et 48,2 % des femmes).

L'AVC est la première cause de mortalité chez la femme avant le cancer du sein et **la troisième chez l'homme**.

Il est la première cause de handicap acquis de l'adulte et la deuxième cause de démence après la maladie d'Alzheimer.



Prévenir que guérir



Ayant une appétence particulière pour le domaine de la santé, cette problématique a particulièrement attisé votre intérêt. Vous décidez de développer un outil d'aide à la prévention d'AVC. Vous accédez au dataset accessible [ici](#).

Cet ensemble de données est utilisé pour prédire si un patient est susceptible d'être victime d'un accident vasculaire cérébral en fonction de paramètres d'entrée tels que le sexe, l'âge, diverses maladies et le tabagisme. Chaque ligne des données fournit des informations pertinentes sur le patient. Le dataset comporte les caractéristiques suivantes :

- **id** : identifiant unique.
- **gender** : genre du patient.
- **age** : âge du patient.
- **hypertension** : 0 si le patient n'a pas d'hypertension, 1 si le patient a de l'hypertension.
- **heart_disease** : 0 si le patient n'a pas de maladie cardiaque, 1 si le patient a une maladie cardiaque.
- **ever_married** : si le patient a déjà été marié ou pas.
- **work_type** : type de travail du patient.
- **Residence_type** : type de résidence du patient "Rural" ou "Urbain".
- **avg_glucose_level** : taux moyen de glucose dans le sang.



- **bmi** : indice de masse corporelle.
- **smoking_status** : "a déjà fumé", "n'a jamais fumé", "fumé" ou "Inconnu".
- **stroke** : 1 si le patient a eu un accident vasculaire cérébral ou 0 s'il n'en a pas eu.

Vous suivez les étapes suivantes afin de réaliser votre projet :

1. Tout bon projet en science des données débute avec un bon prétraitement de données. Nettoyez vos données et rendez-les exploitables.
2. Analysez l'ordre de grandeur et l'unité des différentes données, qu'allez-vous faire ?
3. Une bonne exploitation de données est souvent plus utile que n'importe quelle modélisation de données. Extrayez un maximum d'informations et notez-les. Cela vous sera utile lors des prochaines étapes.
4. Réalisez une veille sur la classification supervisée et sur au moins 3 de ses algorithmes.
5. Modélisez vos données à l'aide de ces trois algorithmes, utilisez une seule feature, la plus intéressante pour vous.
6. Évaluez les résultats de votre classification binaire. Utilisez plusieurs métriques d'évaluation de modèles de classification supervisée et justifiez votre choix.
7. Etes-vous satisfait des résultats obtenus ? Non ? Reprenez votre modélisation de données avec plus de features. Justifiez votre choix avec la sélection de features.



8. Comparez les résultats obtenus précédemment et concluez sur la modélisation la plus efficace pour la prévention des attaques cérébrales. Quels sont les facteurs qui contribuent aux risques ?
9. (BONUS) Déployez votre solution à l'aide de l'outil de votre choix (par exemple Flask ou Dash).

Compétences visées

- Analyse de données
- Apprentissage automatique

Rendu

L'évaluation de ce projet se fera sur deux aspects :

1. Une présentation explicative de votre travail sous forme de diapositives.
2. Un repository github public nommé **stroke-alert**, contenant les éléments suivants:
 - a. Un **notebook Python propre et commenté** (introduction, conclusion, etc) contenant le procédé de développement de votre outil, du nettoyage à la modélisation des données, en passant par l'**analyse exploratoire**. **Pensez à répondre à la problématique**. Vous pouvez avoir au maximum deux notebooks, un pour l'exploration et l'autre pour la modélisation de données.
 - b. Un fichier **README.md** présentant le contexte du projet, les données et leur analyse, les algorithmes utilisés et une conclusion sur votre travail. Pensez à inclure la veille réalisée.
- (BONUS) Le script **.py** de votre solution déployée.



Base de connaissances

- [A Gentle Introduction to Machine Learning](#)
- [5 Types of Classification Algorithms in Machine Learning](#)
- [Classification Algorithms Comparaison](#)
- [Logistic Regression](#)
- [Machine Learning Fundamentals: Cross Validation](#)
- [Metrics to Evaluate your Classification Model to take the right decisions](#)
- [Flask - Quickstart](#)