

Master : Systèmes d'Information Décisionnels et Imagerie

Module : Analyse de données

Année universitaire : 2022/2023

Projet sous le thème :

Analyse de données de santé : Prédiction des accidents vasculaires cérébraux



Réalisé par:

Sihame **BOUTAHRI**

Omayma **SENHOU**

Hassan **RAFYA**

Hamza **ARZIKI**

Oussama **LIMANI**

Responsable:

Pr. Saleh **BOUARAFA**

Table de figures :

Figure 1:Visualisation des variables .	5
Figure 2:Distribution des chaque variable.	5
Figure 3:Variable « Stroke » en fonction des autres variables on utilisant « boxplot »	6
Figure 4: Variable « Stroke » en fonction des autres variables on utilisant « histplot »	6
Figure 5:Visualisation de variable « stroke » en fonction d' Age	7
Figure 6:Matrice de corrélation .	8
Figure 7:la première page d'Interface pour le login.	9
Figure 8:La deuxième page de l'interface.	10

I. Introduction générale :

Un accident vasculaire cérébral (AVC), également connu sous le nom de "stroke" en anglais, se produit lorsque l'approvisionnement en sang vers une partie du cerveau est interrompu.

L'AVC est une urgence médicale et nécessite une prise en charge immédiate. Le traitement dépend du type d'AVC et peut inclure l'administration de médicaments pour dissoudre les caillots sanguins, la chirurgie pour réparer les vaisseaux sanguins endommagés ou l'élimination du sang accumulé dans le cerveau. La rééducation et la réadaptation sont également importantes pour aider les personnes à se rétablir après un AVC.

Il est important de prendre des mesures pour réduire les risques d'AVC, comme maintenir une tension artérielle saine, contrôler le diabète, maintenir un poids santé, faire de l'exercice régulièrement, éviter le tabagisme et limiter la consommation d'alcool. Si vous ou quelqu'un d'autre présente des symptômes d'un AVC, il est essentiel de demander une aide médicale d'urgence immédiatement.

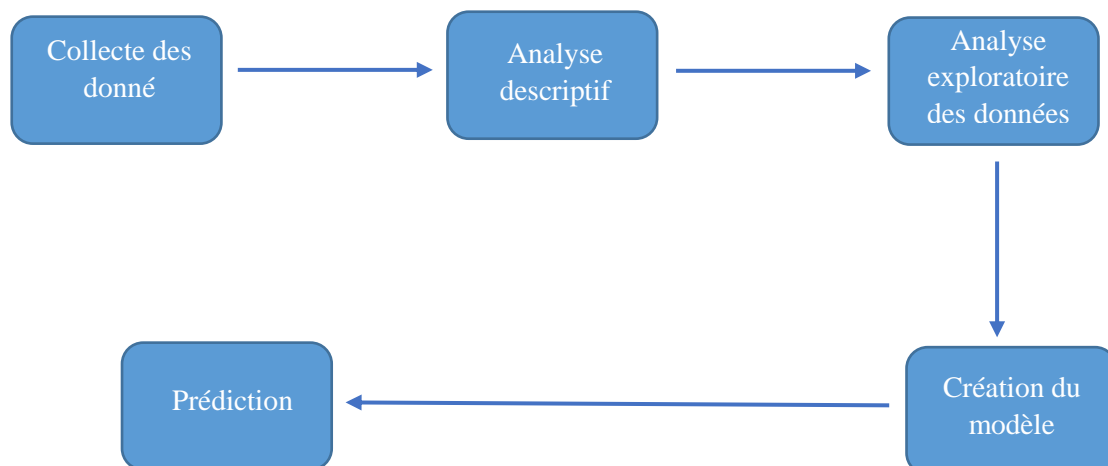
Pour cela nous avons utilisé l'analyse de données pour contribuer à la compréhension, à la prévention et à la gestion des AVC en fournissant des informations précieuses sur les facteurs de risque, les schémas de présentation, les réponses au traitement et les résultats à long terme.

L'objectif :

L'objectif de cette étude est de prédire les accidents vasculaires cérébraux (AVC) en utilisant des données de santé. En utilisant un modèle d'apprentissage automatique, nous cherchons à identifier les facteurs qui sont associés à un risque accru d'AVC et à créer un modèle qui peut prédire si une personne est susceptible de faire un AVC ou non.

En analysant des caractéristiques telles que le genre, l'âge, la présence d'hypertension artérielle, les antécédents de maladies cardiaques, le statut matrimonial, le type de travail, le lieu de résidence, le taux moyen de glucose dans le sang, l'indice de masse corporelle (IMC) et le statut tabagique, nous cherchons à trouver des relations ou des schémas qui peuvent aider à prédire les risques d'AVC.

Une fois que le modèle est entraîné sur un ensemble de données d'apprentissage, il peut être utilisé pour faire des prédictions sur de nouvelles données afin d'estimer la probabilité qu'une personne donnée soit sujette à un AVC.



II. Collecte des données :

La collecte des données a été réalisée à partir de diverses sources, notamment des bases de données médicales, des dossiers électroniques et des registres d'AVC. Les critères de collecte ont été soigneusement définis, en tenant compte de l'âge, du genre, des antécédents médicaux et d'autres variables pertinentes. L'autorisation éthique a été obtenue pour garantir la confidentialité et la protection des informations personnelles des participants. Les données collectées ont ensuite été soumises à un processus de nettoyage et de prétraitement afin d'éliminer les valeurs aberrantes et les données manquantes, assurant ainsi la qualité et l'intégrité des données analysées. Cette étape a été cruciale pour obtenir des résultats fiables et significatifs. Les résultats de l'analyse ont permis d'identifier des facteurs de risque significatifs tels que l'hypertension, les antécédents de maladies cardiaques et le tabagisme, qui sont fortement associés aux AVC. De plus, un modèle de prédiction basé sur l'apprentissage automatique a été développé, montrant une précision prometteuse dans l'identification précoce des individus à risque élevé d'AVC. Ces résultats soulignent l'importance de la prévention et de la prise en charge précoce des facteurs de risque afin de réduire l'incidence des AVC et d'améliorer les résultats cliniques pour les patients.

III. Analyse descriptif :

L'analyse descriptive est une étape importante dans l'étude des accidents vasculaires cérébraux. Elle permet de fournir une vue d'ensemble des données collectées et de comprendre les caractéristiques clés des individus touchés par cette condition médicale.

Cette analyse descriptive initiale fournit une première compréhension des caractéristiques démographiques et des facteurs de risque associés aux accidents vasculaires cérébraux. Ces informations serviront de base pour des analyses plus avancées, telles que l'exploration des relations entre les variables, l'évaluation des modèles prédictifs et l'identification des interventions préventives potentielles.

IV. Analyse exploratoire des données :

L'analyse exploratoire des données est une étape cruciale dans toute étude sur les accidents vasculaires cérébraux. Elle permet de comprendre les caractéristiques des données, d'identifier les schémas et les tendances, et de formuler des hypothèses pour des analyses plus approfondies.

Dans cette étape nous avons fait la visualisation des données pour connaître les caractéristiques de chaque variable. Voici les différentes visualisations que nous avons faites :

Visualisation de chaque variable :

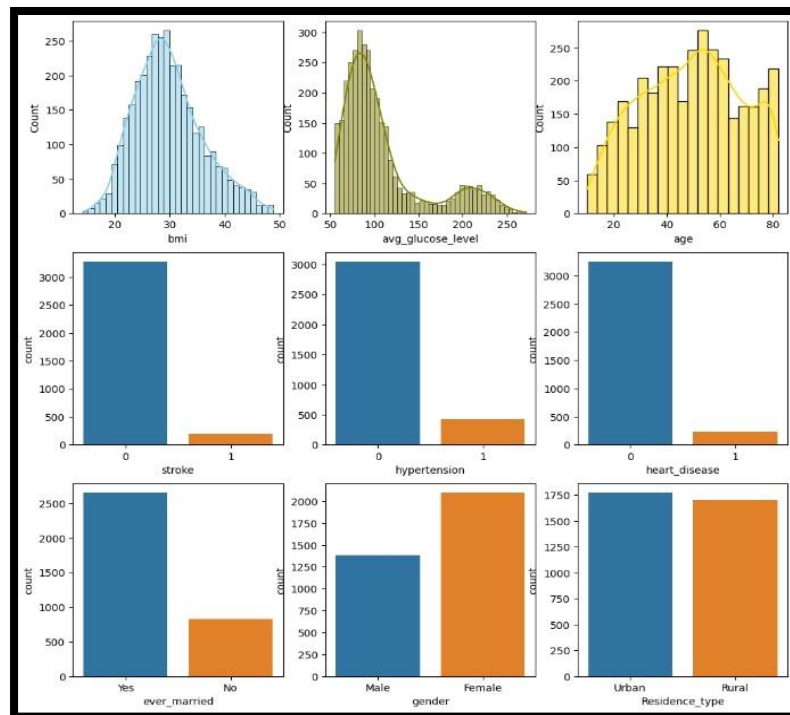


Figure 1: Visualisation des variables .

Interprétation des résultats :

Dans les figures on observe que les patients qu'avaient un accident vasculaire cérébraux est moins que les autres. et concernant les gens mariés on observe que les gens mariés est plus que les autres . Aussi les femmes et plus que les hommes.

+ Distribution de chaque variable en fonction des autres :

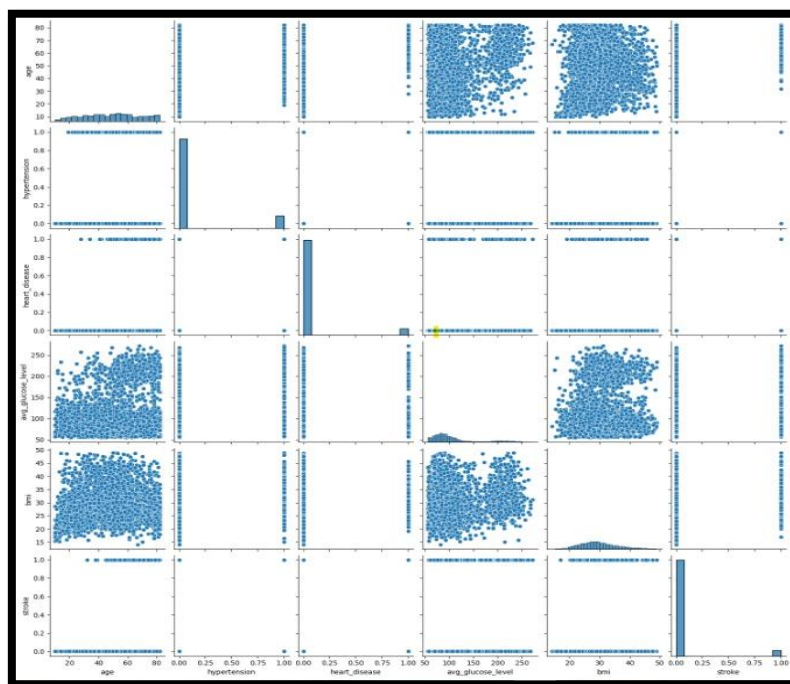


Figure 2: Distribution des chaque variable.

Variation de variable « stroke » en fonction des autres variables :

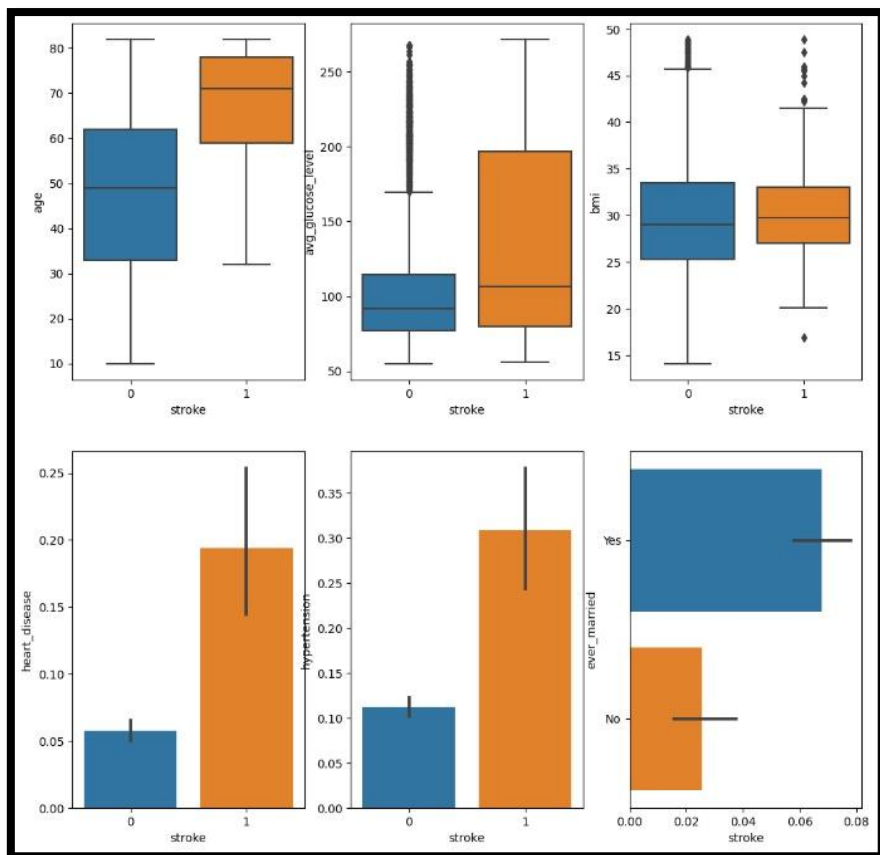


Figure 3: Variable « Stroke » en fonction des autres variables on utilisant « boxplot » .

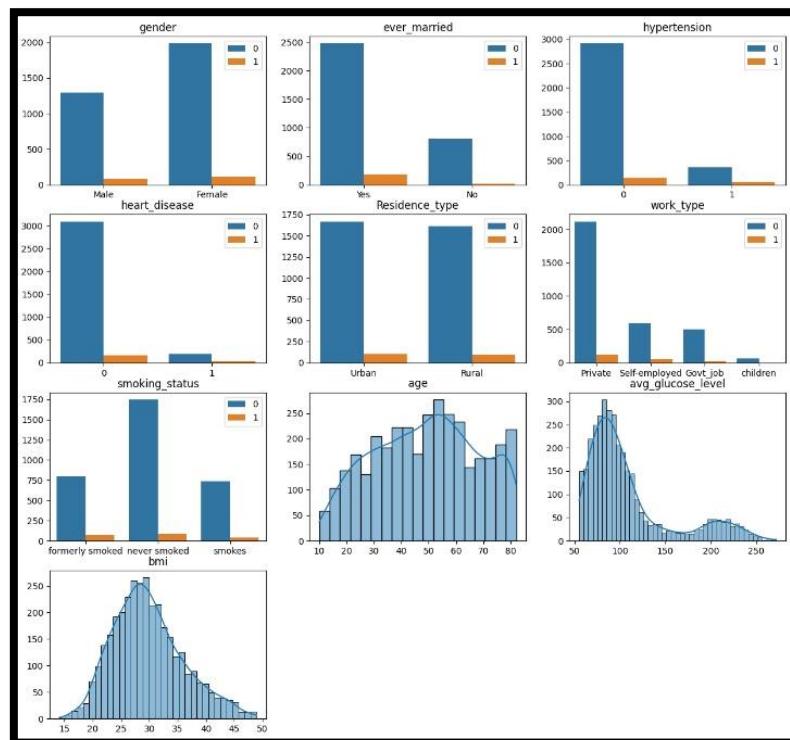


Figure 4: Variable « Stroke » en fonction des autres variables on utilisant « histplot » .

Interprétation des résultats :

Les figures ci-dessus montre la variation de « stroke » en fonction des autres variables . On observe que les femmes sont plus exposées que les hommes. Aussi on a les gens mariés sont plus exposées que les non mariés. On trouve aussi que les gens qui travaillent dans des sociétés privées sont les plus exposées que les gens qui fait des travaux libres...on trouve également que les gens fument sont les moins exposés que les autres.

+ Visualisation plus claire de variable « stroke » en fonction de variable « Age » :

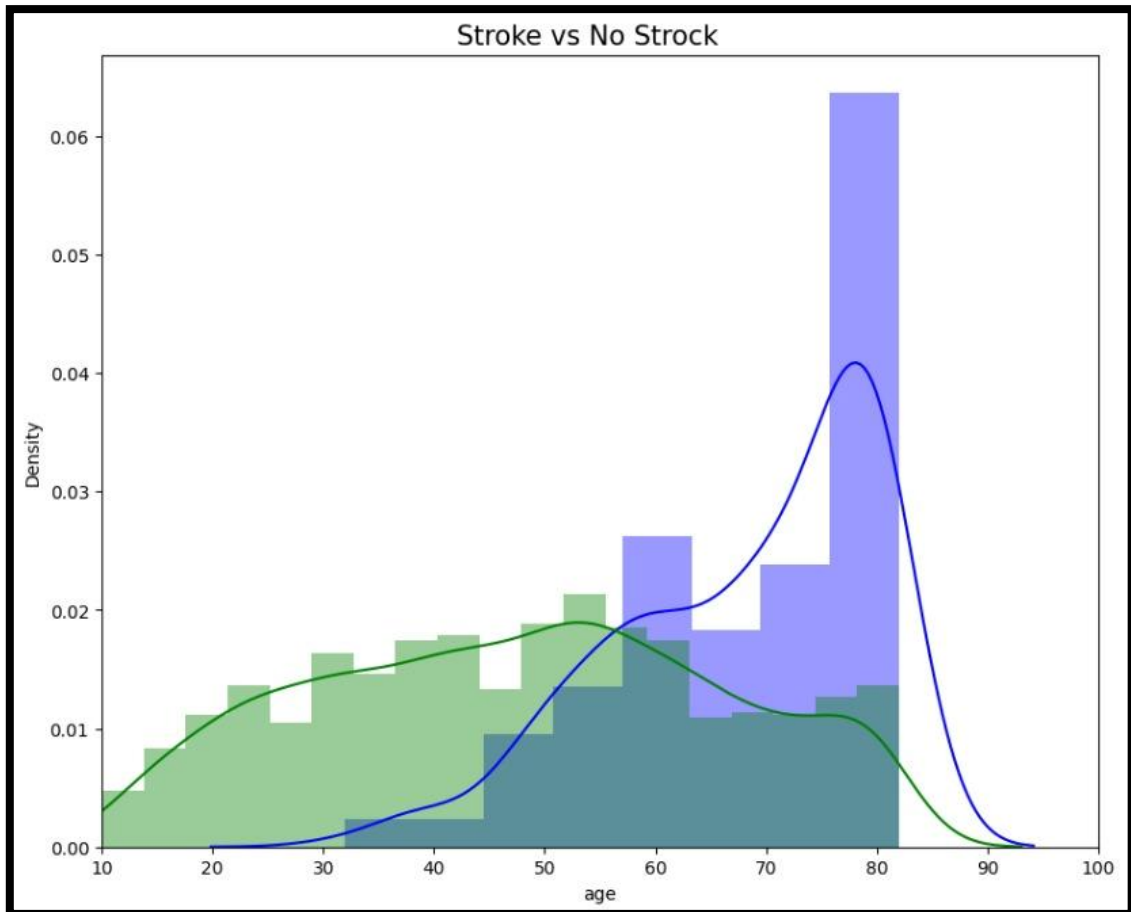


Figure 5: Visualisation de variable « stroke » en fonction d'Age .

Interprétation des résultats :

La figure ci-dessus montre la variation de la densité d'accident vasculaire cérébraux, on observe que les gens âgés de [58-82] sont les plus exposés au accident vasculaire cérébraux que les autres.

La matrice des corrélations des modalités :

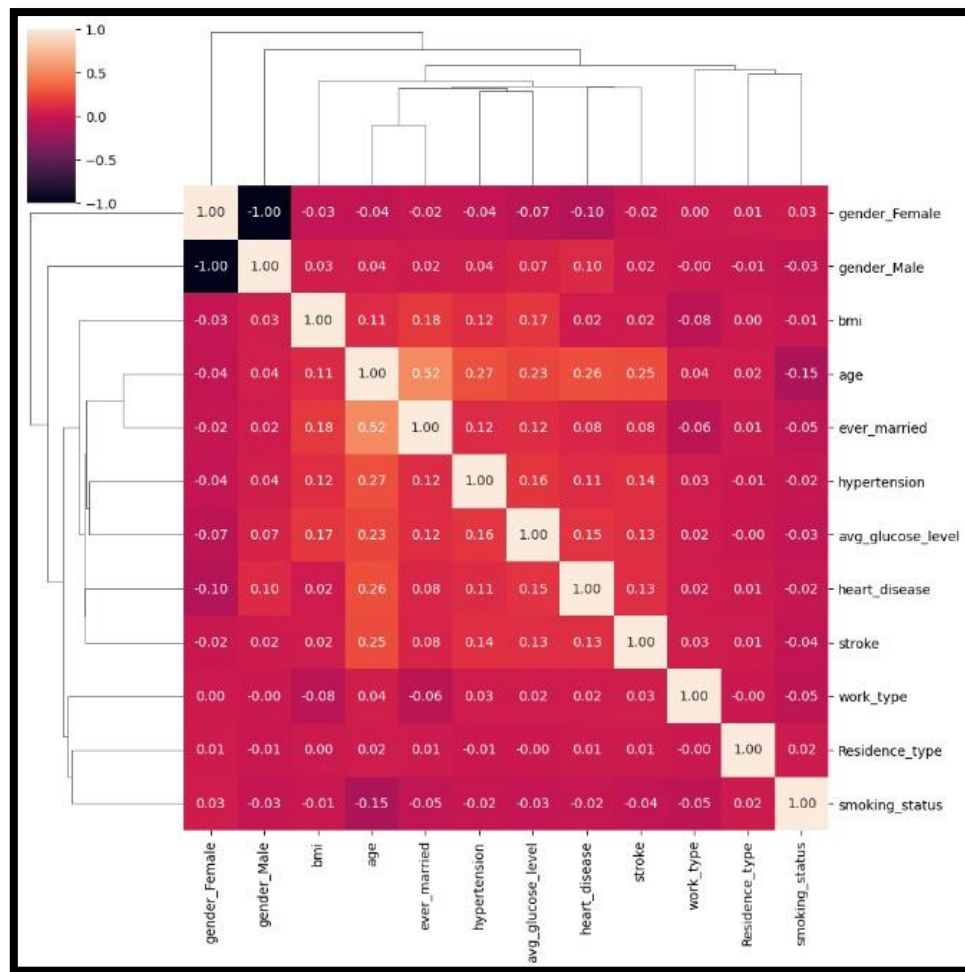


Figure 6:Matrice de corrélation .

V. Création du modèle :

Dans cette étape on a utilisé l'algorithme des K plus proches voisins ou K-nearest neighbors (kNN), c'est un discriminant d'apprentissage supervisé non paramétrique, qui utilise la proximité pour effectuer des classifications ou des prédictions sur le regroupement d'un point de données individuel. Il est généralement utilisé comme algorithme de classification, en partant de l'hypothèse que des points similaires peuvent être trouvés les uns à côté des autres.

L'algorithme K-nn s'effectue en six étapes:

Étape 1 : Sélectionner le nombre K de voisins

Étape 2 : Calculer de la distance (Manhattan/Euclidienne).

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

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

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

Étape 6 : Le modèle est prêt.

VI. Prédiction :

La prédiction de classe pour la nouvelle instance basée sur le modèle KNN entraîné basé sur les étapes suivant :

- ✓ Créer une instance de la classe KNeighborsClassifier en spécifiant le nombre de voisins (k) que vous souhaitez considérer pour la prédiction.



```
Précision du modèle : 0.8996478873239436
```

- ✓ Entraîner le modèle KNN sur vos données d'entraînement en utilisant la méthode fit ().
- ✓ Prévoir une nouvelle instance : Vous pouvez maintenant effectuer une prédiction sur une nouvelle instance en utilisant la méthode predict(). Assurez-vous que la nouvelle instance est prétraitée de la même manière que vos données d'entraînement.

Pour réaliser cette étape nous avons créé une interface pour que les utilisateurs puissent prévoir ses états concernant les accidents vasculaires cérébraux à l'avenir.

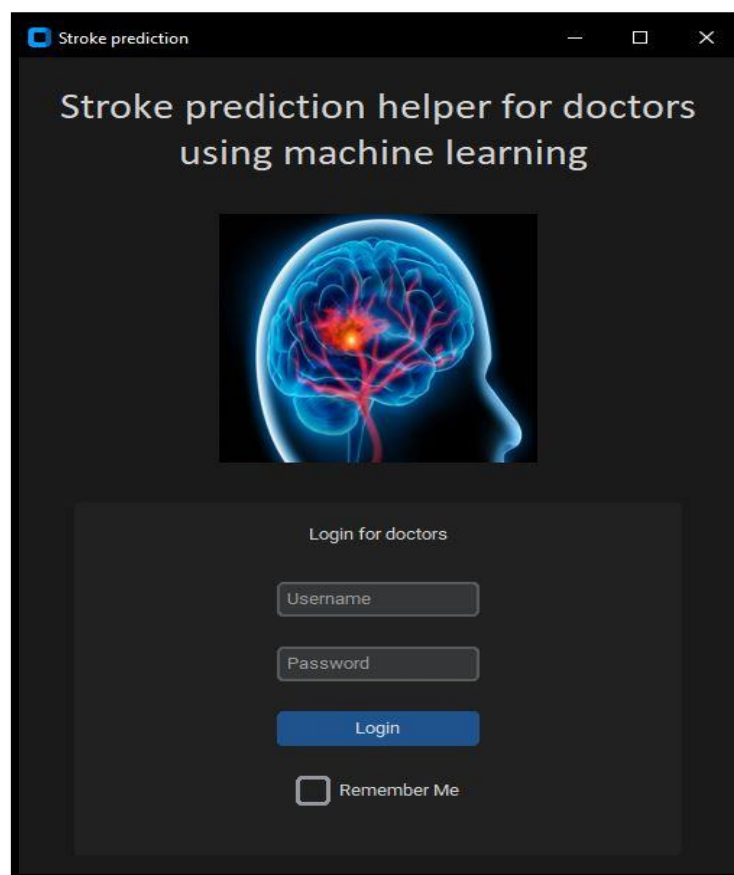


Figure 7: la première page d'Interface pour le login.

La première interface c'est pour le login, l'utilisateur doit saisir son nom et son mot de passe afin d'accéder à la deuxième page pour faire la prédiction d'un nouveau patient.

Figure 8: La deuxième page de l'interface.

Alors dans cette page le médecin ou l'utilisateur en générale saisit toutes les informations de patient ou de lui-même pour prévoir son état concernant le problème de l'accident vasculaire cérébral, donc il faut saisir le nom, l'âge, le statut, le BMI, le niveau de glucose, le type de travail, le type de résidence..... Afin d'afficher les résultats.

VII. Conclusion :

Il y a un total de 8 informations trouvées dans l'ensemble de données sur l'AVC :

- 1) Il semblait que l'IMC et l'âge étaient positivement corrélés, bien que l'association ne soit pas forte.
- 2) Le patient plus âgé était plus susceptible de subir un accident vasculaire cérébral qu'un patient plus jeune.
- 3) Un IMC plus élevé n'augmente pas le risque d'AVC.
- 4) Le diabète est l'un des facteurs de risque d'AVC et les patients prédiabétiques ont un risque accru d'AVC.
- 5) Une proportion plus élevée de patients souffrant d'hypertension ou de maladie cardiaque ont subi un accident vasculaire cérébral, toutes choses étant égales par ailleurs.
- 6) Peu importe le sexe du patient et l'endroit où il est resté, il a la même probabilité de subir un AVC.
- 7) La variable du type de travail était fortement associée à l'âge.
- 8) La variable relative à l'état matrimonial était fortement associée à l'âge.

Références :

1. https://rpubs.com/yujie_deng/766152
2. <https://towardsdatascience.com/step-by-step-exploratory-data-analysis-on-stroke-dataset-840aefea8739>
3. <https://medium.com/geekculture/performing-exploratory-data-analysis-on-stroke-dataset-1b4885989030>
4. <https://www.kaggle.com/>