**Prédiction du risque d’avoir un AVC avec le modèle de Random Forest**

**Description du problème** : Selon l’Organisation Mondiale de la Santé (OMS) Chaque année, 15 millions de personnes font un accident vasculaire cérébral (AVC) : 5 millions d'entre elles meurent et 5 millions souffrent d'une incapacité permanente, ce qui représente un poids pour la famille et la communauté. La suspicion du diagnostic d'AVC repose en règle générale sur la clinique avec deux éléments clés à savoir un déficit neurologique focalisé et une apparition brutale. L'examen neurologique confirme le déficit, en précise la topographie et permet d'évoquer le territoire atteint. C’est dans le souci d’éviter les erreurs de diagnostic et de coût lié aux examens que nous proposons un modèle de Machine Learning qui détectera le risque d’un risque d’avoir un AVC en fonction de certaines caractéristiques.

**Outil :** Dans notre recherche, nous avons utilisé le langage de programmation python et le Forêts d'arbres décisionnels (Random Forest) comme algorithme pour traité le sujet. Nous avons utilisé les bibliothèques : pandas, numpy, seaborn, matplotlib, sklearn,

**Données** : Nous avons utilisé les données de Kaggle et sont disponibles sur notre espace de repos (<https://github.com/Sogo95/Pr-diction-du-risque-d-AVC>)

**Méthodologie :**

* Nous avons procédé à la connaissance des données par leur exploration.
* Nous avons traité les variables manquantes, encoder et normalisé certaines de nos variables explicatives.
* Nous avons observé la distribution de la variable d’intérêt (risque d’avoir un AVC) et avons fait ressortir les variables qui étaient corrélé au risque d’AVC ainsi que celles ayant un effet sur cette même variable.
* Les variables asymétriques ont été également détecté et corriger
* Nous avons procédé à la construction du modèle tout en divisant notre base de données en ensemble d’entrainement (70%) et de test (30%).
* Le Random Forest a été utilisé et une prédiction sur l’ensemble des données test a été effectué.
* Nous avons terminé par l’évaluation de notre modèle et avions trouvé un accuracy de 99,6%. Ainsi notre modèle permet de prédire correctement le risque d’avoir un AVC à 99,6%