

TEST TECHNIQUE

EXPLORATION DE DONNEES

Le test est à réaliser dans un **notebook jupyter python** qui sera soit transmis dans un repository git, soit hébergé sur un service au choix du candidat.

Le notebook devra contenir les explications et informations nécessaires à la compréhension et à la vérification.

Le test est à réaliser en 1h. Si vous manquez de temps pour tout faire, ce n'est pas grave, ce sera l'occasion de poursuivre la discussion dans la suite du processus.

Dataset

Le set de données est composé des 10 colonnes suivantes et 10 000 individus:

1. id_individu : id de l'individu
2. age : age de l'individu
3. ic_atcd : antécédent d'insuffisance cardiaque dans les 3 ans précédant l'hospitalisation
4. hypertension_atcd : antécédent d'hypertension dans les 3 ans précédant l'hospitalisation
5. aoc_atcd : prise d'anti-coagulants dans les 3 mois précédant l'hospitalisation
6. aoc_post_hosp : prise d'anti-coagulants dans les 3 mois suivant l'hospitalisation
7. pacemaker_hist : antécédent de pacemaker
8. cardio_consultation_hist : antécédent de consultation avec un cardiologue dans les 3 mois précédant l'hospitalisation
9. complication_intra_hosp : complication au cours de l'hospitalisation
10. mortalite_intra_hosp : décès au cours de l'hospitalisation

Questions

1. Combien d'individus sont âgés de plus de 60 ans et présentent un antécédent d'insuffisance cardiaque ou d'hypertension ?
2. Quelle est la distribution d'âge des individus de moins de 60 ans ?
3. Y-a-t-il une différence statistiquement significative d'avoir un antécédent d'insuffisance cardiaque entre les individus de plus et de moins de 60 ans ?
4. Y-a-t-il une différence statistiquement significative d'âge entre les individus ayant ou non un antécédent d'insuffisance cardiaque ?

5. A l'aide de régressions logistiques univariées et multivariées, quelles variables du set de données permettent d'expliquer la survenue d'une complication hospitalière ? La méthodologie à utiliser est la suivante : les variables doivent être sélectionnées sur la base des résultats des modèles univariés avec une significativité supérieure à 0.2.
6. (Bonus) A l'aide du modèle de Machine Learning de votre choix, décrivez la démarche que vous auriez adoptée, et les outils python que vous auriez utilisés pour prédire la mortalité intra hospitalière (décès au cours d'une hospitalisation).

Tips : n'hésitez pas à explorer le set de données.