

## Machine dans le domaine de l'assurance maladie

### CONTEXTE

Une entreprise évoluant dans le domaine de l'assurance maladie à sa disposition des données relatives à la consommation annuelle en frais médicaux de ses clients. L'entreprise voudrait tirer profit de ces données en les analysant et en développant à partir celles-ci un modèle de Machine Learning permettant de prédire avec le minimum d'erreurs possibles la consommation annuelle de ses clients. Ceci lui évitera de sous-évaluer ou surévaluer les primes facturées à ses clients. Ceci aura l'avantage d'éviter des pertes financières à l'entreprise en cas de sous-évaluation des primes et des pertes de clients en cas de surévaluation des primes.

### Description du jeu de données

Le jeu de données mis à votre disposition pour faire cette analyse est constitué de certaines caractéristiques pertinentes des clients :

- **age** : l'âge du client ou de l'assuré principal.
- **sex** : le sexe de l'assuré principal.
- **bmi** : l'indice de masse corporelle du client.
- **children** : le nombre d'enfant de l'assuré principal.
- **smoker** : si le client est un fumeur ou pas.
- **region** : la région où habite le client.
- **charges** : la consommation annuelle du client en frais médicaux.

### TRAVAIL A FAIRE

Avec "**charges**" comme variable cible et "**age**" et "**bmi**" comme variables indépendantes et après avoir analysé la corrélation entre la variable cible et les variables indépendantes, vous développerez un modèle de régression linéaire pour prédire la variable cible. Le modèle sera entraîné d'abord avec chacune des variables indépendantes séparément avant de l'entraîner en utilisant simultanément les deux variables indépendantes. Vous évaluerez à chaque le modèle entraîné en utilisant les métriques appropriées.

Date de présentation : 15 février 2024