# **Tp Qualité du vin**

Contexte :

Votre client est un producteur de vin. Il aimerait pouvoir identifier la qualité du vin qu’il est en train de produire à partir de différents critères chimiques. Il dispose pour cela d’un jeu de données, où pour plusieurs bouteilles de vins où le contenu a été analysé chimiquement et une note de qualité de vin a été associée. Votre objectif est d’automatiser le processus de qualification du vin

Objectif :

Réfléchissez à une méthodologie complète pour construire un modèle capable de prédire pour une bouteille de vin disposant de caractéristiques chimiques quelle qualité le vin pourrait prétendre. Vous devrez traiter ce problème comme un problème de **classification.**

Votre méthodologie devra répondre aux questions suivantes :

* Comment donner confiance à votre modèle ? Comment avoir une idée de performance de votre modèle ?
* La distribution de mes classes (notes de qualité) est-elle égale ?

Conseil des étapes de résolution du problème :

* Comprendre le jeu de données (statistiques, visualisations ?)
* Sélectionnez un premier modèle de classification, implémentez-le
* Entrainer votre modèle
* Evaluer votre modèle
* DIAGNOSTIQUEZ votre modèle (under/overfitting ? une classe que vous prédisez moins bien ? etc.)
* Itérer et recommencez 😊

Conseil :

L’étape la plus importante à comprendre à ce stade réside dans l’évaluation du modèle. Cette étape est souvent négligée et pourtant il est primordial de la comprendre. Si ce n’est pas clair demandez moi 😊. Ensuite, l’étape de diagnostique et également oubliée, attention si vous entrainez votre modèle ce n’est pas la fin, mais le début du projet !

Faites une première itération seul(e), puis n’hésitez pas à échangez avec une autre personne pour voir comment ils ont aborder le problème.

Librairies python conseillées :

Sklearn  
pandas  
matplotlib / seaborn

Juputer notebook