

VINHO VERDE

Les données

Le dataset suivant recense un certain nombre de vins portugais. Pour chaque vin, 11 variable quantitatives sont fournies. Vous trouverez la description des variables ici :

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Le dataset est ici : <http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv>

L'objectif

Nous souhaitons, à partir de ces données, mieux comprendre la relation entre la note d'un vin (celle accordé par les usagers - entre 0 et 10) et les différents attributs du vin.

Comme nous l'avons vu en formation, créez un notebook dans lequel vous suivrez les grandes étapes suivantes :

- Analyse des données, calcul de corrélations (**data.corr()** pour avoir les corrélations des différentes variables entre elles)
- Présentation graphique des relations entre les quelques variables les plus importantes et la note du vin (**plt.scatter(x,y)** pour créer un nuage de points à partir de x et y - le pH et la qualité par exemple)
- Calibration d'un premier modèle. Utilisez la « carte des modèles » proposé par sklearn et surtout, surtout la doc sklearn qui recense tout les modèles possibles.

La trame donnée n'est là que pour vous aider, vous êtes libre de faire autrement.

N'oubliez pas l'objectif : prédire la qualité du vin et rien d'autre. Si votre modèle doit se résumer à une structure conditionnelle, qu'à cela ne tienne. Concentrez vous sur l'objectif : comprendre et modéliser les relations entre les attributs quantitatifs du vin et sa note. Si vous faites un super modèle (au sens de la précision) qui est inutilisable, vous ne remplissez pas l'objectif. A l'inverse, si votre modèle est très simple mais utile et exploitable, vous aurez mieux atteint votre objectif.

Je vous invite aussi à échanger entre vous et à me solliciter sur Slack si vous avez des questions. Ne restez pas bloqués sur un souci de syntaxe ou un obstacle technique - sollicitez moi.

<http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv>