

Bilan du projet Python pour un Data Scientist - Nicolas Toussaint

Ce fut mon premier projet de machine learning. J'ai donc pu appréhender pour la première fois les différents modules de Sklearn. Ce fut intéressant de découvrir les nombreuses fonctionnalités de cette librairie. La librairie étant très user-friendly avec des méthodes similaires pour les différents classifieurs. Je me suis rendu compte qu'il était finalement très simple d'avoir un résultat sur une base de données du moment qu'elle est bien structurée. La valeur de ce projet n'est pour moi pas dans le score final que nous réussissons à obtenir mais dans le regroupement des nombreux tests que nous avons effectués. J'ai codé les classes permettant de tester différents classifieurs rapidement et je suis sûr qu'elles me seront vite utiles pour d'autres projets.

Une fois cette prise en main des fonctionnalités de sklearn, j'ai tenté d'en approfondir l'utilisation en créant mon propre classifieur. Ce fut intéressant d'essayer de regrouper les différents résultats obtenus pour créer un classifieur spécifique à notre base de données. Si les résultats ne sont pas probants, ce classifieur montre l'utilité des classes de tests. Elles ont pu tout aussi facilement qu'avec un classifieur de sklearn obtenir des résultats. Cela m'a donc conforté dans l'idée que ces classes me seront profitables dans mes futurs projets.

Au début du semestre, notre projet ne portait pas sur les vins mais sur de l'analyse d'images. Je possède une carte graphique assez puissante (GTX 1070) et je m'étais donc lancé dans l'idée d'utiliser Keras avec ma carte graphique. Malheureusement, après une matinée à essayé d'installer les librairies sous Linux, réinstaller plusieurs fois l'OS après des installations bugées, j'ai dû me résigner. Ce fut une grande déception. Cependant les vastes possibilités de sklearn pour obtenir des résultats m'ont vite rassurées.

Cette première approche du machine learning fut très enrichissante. Mon seul reproche est que notre base de données était à l'origine déjà très bien structurée. Elle n'avait donc pas besoin de traitement supplémentaire. C'est pourquoi dans de futurs projets, j'aimerais m'attarder sur l'acquisition des données et la création de mes propres features. C'est une facette que nous n'avons pu aborder et qui me semble tout aussi passionnante.