Un projet de bout en bout

# End-to-End Machine Learning Project

Nous vous demandons de lire le chapitre 2 du livre suivant

Hands-On Machine Learning with Scikit-Learn & TensorFlow, Aurélien Géron, 2017, p. 33 à 79

Du code python accompagne ce chapitre (02\_end\_to\_end\_machine\_learning\_project.ipynb), il est disponible sous la forme d’un “notebook Jupyter”[[1]](#footnote-1). Vous pouvez soit l’exécuter avec Anaconda[[2]](#footnote-2) ou en ligne avec Google Colab[[3]](#footnote-3)

# Travail à remettre

Pour le mardi 24/09/2019 à 23h59, nous vous demandons de poster sur mooVin les deux choses suivantes :

1. Un résumé d’une page (cad une face A4) du chapitre 2. Ce résumé doit se focaliser sur les 8 étapes d’un projet de machine learning.

Notez que votre résumé peut être manuscrit, dans ce cas, postez une photo de celui-ci sur mooVin

1. Un schéma orignal illustrant le chapitre 2.

# Pour aller plus loin

Cette section contient quelques références qui vous permettront d'aller un peu plus loin que le travail obligatoire de cette semaine. Nous vous suggérons de parcourir ces références et de les garder dans un coin de votre tête. Elles vous seront certainement utiles un jour ou l'autre.

## The Machine Learning Landscape

Le chapitre 1 du livre (p. 3 à 32) pourrait vous intéresser. Il présente une vue générale du "Machine Learning" et ses différentes facettes.

## Python

A Whirlwind Tour of Python is a fast-paced introduction to essential features of the Python language, aimed at researchers and developers who are already familiar with programming in another language. The material is particularly designed for those who wish to use Python for data science and/or scientific programming.[[4]](#footnote-4)

<https://jakevdp.github.io/WhirlwindTourOfPython/>

## Numpy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

* a powerful N-dimensional array object
* sophisticated (broadcasting) functions
* tools for integrating C/C++ and Fortran code
* useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.[[5]](#footnote-5)

<https://docs.scipy.org/doc/numpy/user/quickstart.html>

## Pandas

*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](https://www.python.org/) programming language.[[6]](#footnote-6)

<https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html>

La librairie pandas est à mi-chemin entre une base de données relationnelle est une feuille Excel. Le tout en Python. Cette librairie peut être très utilise pour visualiser ou présenter des données. Typiquement au début ou à la fin d'un projet de Machine Learning.

1. <https://jupyter.org/> [↑](#footnote-ref-1)
2. [www.anaconda.com](https://hevinci.sharepoint.com/sites/I3100-IntelligenceArtificielle/Documents%20partages/General/Préparations/S2/www.anaconda.com) [↑](#footnote-ref-2)
3. [https://colab.research.google.com](https://colab.research.google.com/) [↑](#footnote-ref-3)
4. <https://pandas.pydata.org/> [↑](#footnote-ref-4)
5. [https://numpy.org](https://numpy.org/) [↑](#footnote-ref-5)
6. <https://pandas.pydata.org/> [↑](#footnote-ref-6)