[Curso de Machine Learning Aplicado con Python](https://platzi.com/clases/scikit/)

**ArtículoMontar un ambiente de trabajo de Machine Learning**

[**Juan Pablo Morales**](https://platzi.com/%40juan-pablo-morales)

19 de Marzo de 2019

**Montar un ambiente de trabajo de Machine Learning**

Existen diversas opciones para montar un ambiente de trabajo de Machine Learning. De forma general lo que queremos es poder tener python 3, jupyter notebooks y las librerias esenciales del ambiente Pydata:

* Numpy
* Pandas
* Matplotlib
* Scikit-learn

Dependiendo de nuestro objetivo necesitaremos herramientas más simples para estar trabajando rápidamente, o de herramientas más complejas que permiten tanto colaboración como pasos a producción.

En los videos del curso proponemos el uso de Docker, que es una de las herramientas más completas -y la que usamos junto a mi equipo en Arara- por lo que en este post les propondré algunos setups alternativos
que les pueden hacer más rápida la etapa de configuración si es que desean entrar de lleno a testear el código del curso o prototipar modelos.

**Anaconda GUI**

Anaconda provee una plataforma completa de data science con una interfaz gráfica para acceder fácilmente a sus aplicaciones, de las cuales la que nos interesa es Jupyter. Si bien la versión en línea de comando (miniconda) es mejor para compartir ambientes entre programadores, la versión con interfaz gráfica es probablemente una de las opciones más fáciles de instalar en local en un computador.

Para instalarla deben:

* Ir a <https://www.anaconda.com/distribution/>
* Elegir su sistema operativo: Windows/Linux/macOS
* Descargar el **instalador gáfico** para la arquitectura que corresponda a la CPU que posee su computador (32 bits o 64 bits)



Cuando hayan descargado el instalador gráfico y lo hayan abierto se les preguntara diversas confirmaciones, de forma general las opciones por defecto no necesitan modificarlas.


Anaconda quedara luego de estos pasos instalado dentro de sus aplicaciones. Al abrirlo en la home tienen Jupyter, y pueden abrirlo con botón Launch.

El ambiente en el que estarán trabajando es un ambiente de base que define Anaconda, llamado “base (root)” y que contiene las librer�as esenciales de Pydata que les mencionamos.
Sin embargo las versiones de las librerías podrán ser distintas a las definidas en el Docker, por lo que si tuviesen diferencias con los resultados del curso, lo pueden mencionar en los comentarios o preguntas.

**Google Colab**

Si cuentan con una cuenta Google esta es la forma más rápida de tener un ambiente de prototipado. La interfaz es muy parecida a la de Jupyter (solo con algunas pequeñas modificaciones) y crear un nuevo notebook es tan rápido como crear un documento en Google docs. Sin embargo si trabajan con Google Colab no podrán trabajar con archivos locales en su computadora, por lo que el código del curso lo tendrán que modificar en varios videos para poder cargar datos. Una instalación instantánea pero más trabajo para poder seguir el curso.

Para acceder deben ingresar a <https://colab.research.google.com/>, logearse y crear un nuevo notebook de Python 3.



La forma más fácil de cargar datos en Google Colab es desde Github:

* Ir a github del curso:
* Abrir link del archivo que desean importar a colab (por ejemplo si quiero importar “” )
* Clickear en “View Raw”
* Copiar la URL (en el ejemplo de código que sigue la llamaremos ‘url<sub>archivo</sub><sub>raw</sub><sub>copiada</sub>’)
* En el código importar el archivo de la forma siguiente

import pandas as pd

url = ‘url\_archivo\_raw\_copiada’

df = pd.read\_csv(url)

¡Con esto ya tenemos dos opciones además de Docker! También se pueden ver algunas excelentes alternativas en los comentarios de los videos como las sugerencias.

¡Gracias y disfruten del curso!