

# IMPLEMENTADO MACHINE LEARNING

## Práctica 1 – Preprocesamiento y Visualización

### Material de lectura:

- Andreas C. Müller, Sarah Guido. Introduction to Machine Learning with Python. Capítulo 1.
- Stephen Marsland. Machine Learning: An Algorithmic Perspective. Capítulo 1.
- Pandas: <https://pandas.pydata.org/>

- 1- **Tutorial Python.** Desde una terminal de Anaconda ubicarse en el directorio donde se encuentra los archivos con extensión “.ipynb” (notebooks) y escribir: “Jupyter notebook”. Desde el entorno web abrir el tutorial “01\_Tutorial de Python con Jupyter Notebook.ipynb”
- 2- **Tutorial NumPy.** Realizar el mismo procedimiento para el tutorial “02\_Tutorial de Numpy.ipynb”.
- 3- **Análisis de dataset.** El archivo “iris.csv” contiene información de 3 especies de flores. Realice las siguientes operaciones. El archivo “análisis\_datasets.py” puede serle de utilidad.
  - a. Utilice la librería Pandas para abrir el archivo.
  - b. Haga un gráfico con la función `scatter_matrix()` para visualizar todos los pares de variables. ¿Qué conclusiones puede obtener de los gráficos?
  - c. Realice un histograma de todas las variables (`dataFrame.hist()`).
  - d. Analice la matriz de correlación para detectar variables relacionadas linealmente.
  - e. Realice un diagrama de caja. Visualice los cuartiles y los outliers.
  - f. Normalice las variables (cada columna) con z-score y vuelva a analizar los datos.
- 4- **Preprocesamiento.** Analice los siguientes datasets e intente tener una vista minable de los datos. Elimine valores nulos, descarte columnas innecesarias, convierta atributos nominales, analice la correlación entre variables. Puede utilizar las funciones de pandas `describe()`, `isnull()`, `dropna()`, `corr()`, entre otras.
  - a. diabetes.csv
  - b. who\_life\_expectancy.csv
  - c. weatherAUS.csv
  - d. breast\_cancer.csv