A1_DatasetIris

Pamela Sánchez Arellano A01636995

In [1]:
```python
# Define where you are running the code: colab or local
RunInColab            = True      # (False: no  | True: yes)

# If running in colab:
if RunInColab:
    # Mount your google drive in google colab
    from google.colab import drive
    drive.mount('/content/drive')

    # Find location
    #!pwd
    #!ls
    #!ls "/content/drive/My Drive/Colab Notebooks/MachineLearningWithPyt
hon/"

    # Define path del proyecto
    Ruta           = "/content/drive/My Drive/Colab Notebooks/TC1002S/N
otebooksStudents/A01636995"

else:
    # Define path del proyecto
    Ruta           = "/Users/pamelasanchez/Documents/TC1002S/NotebooksS
tudents/A01636995"
```

Mounted at /content/drive

In [3]:
```python
# Import the packages that we will be using
import matplotlib.pyplot as plt
import pandas as pd

# Dataset url
url = Ruta + "/datasets/iris/iris.csv"

# Load the dataset
dataset = pd.read_csv(url)
```

```
In [4]: # Print the dataset
        dataset
```

Out[4]:

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
...	...	...	...	...	...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

149 rows × 5 columns

```
In [7]: # Print the number of rows
        dataset.shape[0]
```

Out[7]: 149

```
In [8]: # Print the number of columns
        dataset.shape[1]
```

Out[8]: 5

Activity: work with the iris dataset

1. Load the iris.csv file in your computer and understand the dataset

1. How many observations (rows) are in total? There are a total of 149 rows.

1. How many variables (columns) are in total? What do they represent? There are a total of 5 columns and they represent the dimensional characteristics of each flower which is represented as a variable.

1. How many observations are for each type of flower? 4
2. What is the type of data for each variable? For the first 4 observations the variables are numerical (float) and the last observations is a string.
3. What are the units of each variable? For the first 4 observations the units are float numbers and the last one, string is characters.

In [15]: `dataset.dtypes`

Out[15]:
```
5.1              float64
3.5              float64
1.4              float64
0.2              float64
Iris-setosa       object
dtype: object
```