Artificial Intelligence: Lab 0

Data Preprocessing

In this first lab, we will introduce some data preprocessing tools. Data preprocessing refers to the steps applied to make data more suitable for machine learning. The main steps we will consider are the following :

1. Load the data, for this, we will mainly use `pandas`.
2. Visualise the data. We will use `matplotlib` and `seaborn`.
3. Handle missing values.
4. Rescale data.

For this lab, you need to download the file `l2_lab0_skeleton.py` available at the address :

`https://www.labri.fr/~zemmari/l2-ai`.

We will first work with the dataset `titanic` available at the address :

`https://www.labri.fr/~zemmari/datasets`.

You can find its complete description at :

`https://www.kaggle.com/c/titanic`.

Complete the file `l2_lab0_skeleton.py` to answer the following questions :

1. Load and get an overview of the data.
2. What can you say about the data ? Are there any missing data ?
3. Write `python` instructions to drop all the rows with at least one missing data. What do you observe ?
4. Write `python` instructions to drop the rows only if all of the values in the row are missing.
5. What do you observe about the column `Cabin` ? Write instructions to drop it.
6. Discuss the other columns with missing values. Choose the right strategy to replace the missing values.