Initial Study of the Dataset

Initial Study of the Dataset

In this section we will work on the basic process of analyzing a dataset to train machine learning models.

On this lesson we will perform an initial study of the dataset.

Lucas Bazilio - Udemy

Dataset

It is a dataset of the **titanic** which aims to predict the survival of the passengers of this ship from its attributes.

We have passenger information such as *age*, *sex*, *cabin*…

Dataset Download

You can download the dataset `titanic.csv` from the downloadable resources of this lesson.

Attributes Information

survival - Survival (0 = No; 1 = Yes)

class - Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)

name - Name

sex - Sex

age - Age

sibsp - Number of Siblings/Spouses Aboard

parch - Number of Parents/Children Aboard

ticket - Ticket Number

Attributes Information

fare - Passenger Fare

cabin - Cabin

embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

boat - Lifeboat (if survived)

body - Body number (if did not survive and body was recovered)

home.dest - home destination

Titanic Dataset Information

You can see all of the information of this dataset in:
https://www.kaggle.com/c/titanic