|  |
| --- |
| Master 1 - ISII |
| Artificial Intelligence |
| Machine Learning Project |

|  |
| --- |
| Team N° 10   * 1 * 2 * 3 * 4 * 5 |

Table of Contents

[1. Introduction 3](#_Toc95914655)

[2. Project Descriptions 3](#_Toc95914656)

[2.1. Regression Problem 3](#_Toc95914657)

[2.2. Classification Problem 3](#_Toc95914658)

[3. Real Estate Valuation 4](#_Toc95914659)

[3.1. Dataset Description 4](#_Toc95914660)

[3.2. Preprocessing 4](#_Toc95914661)

[3.3. Linear Regression Model 4](#_Toc95914662)

[4. Immunotherapy Treatment Result 4](#_Toc95914663)

[4.1. Dataset Description 4](#_Toc95914664)

[4.2. Preprocessing 4](#_Toc95914665)

[4.3. Logistic Regression Model 4](#_Toc95914666)

# Introduction

The project that we had to undertake is based on a subcategory of machine learning and artificial intelligence, which is Supervised Learning. Essentially, supervised learning is when the computer is taught by example. It learns from past data and applies the learning to present data to predict future events. In this case, both input and desired output data provide help to the prediction of future events.

Supervised learning can also be divided into two subcategories, classification and regression. Which are the two sorts of problems that we had to operate.

We had to apply the specific analysis methods to each category, to solve the two problems and train different models capable of predicting the best possible results.

On the other hand, we also had to comply with specific working methods to solve a machine learning problem. The use of a notebook is therefore essential for this, because it is used to present the analysis process step by step by arranging the code, images, text, output etc. in a step-by-step manner. Jupyter Notebook provides a full set of features that makes it one the best components of Python Machine Learning. Another tool that we used, this time for the arrangement and organization of work in general, is Git and GitHub, such as each member of the group made changes on the part of the notebook that corresponded with a specific part of the project (preprocessing, graphic visualization etc.)

In the following parts, we are going to discuss the achieved tasks in each of the two problems, explaining the dataset content, the different steps of preprocessing realized to improve its quality, and testing and visualizing the obtained results produced from the built models.

# Project Descriptions

## Regression Problem

One of the two problems that we had to process is a regression problem. A regression analysis must be considered when we deal with a problem that requires a prediction of a continuous value. Our regression problem is about valuating real estate properties given certain characteristics about them and their surroundings. Basically, the main task is to predict a house price, which is continuous value, considering a group of numerical/categorical features describing this house.

## Classification Problem

The second problem is a classification problem. This is the kind of problems that require the result of the prediction to be a categorical value. In our case it is a binary classification, for the reason that the model can only predict a 0 or 1 value for a given set of feature values concerning an instance. Our classification problem is about predicting if Immunotherapy, when used for wart treatment, produces a positive or negative result on a patient, given different informations about this one. A prediction of “1” being a result of a successful treatment and “0” of an unsuccessful one.

# Real Estate Valuation

## Dataset Description

## Preprocessing

## Linear Regression

# Immunotherapy Treatment Result

## Dataset Description

## Preprocessing

## Logistic Regression