**Estimating financial needs – project overview**

This document provides a brief general explanation of the project undertaken, what has been done, the reasons behind certain decisions, and a simple guide for reading the provided notebooks, which can be a useful addition to the final presentation.

The project's goal is to build a recommendation system for two types of financial products (income and accumulation), targeting clients whose features were provided in the initial dataset. Despite numerous attempts, for clarity, we condensed the most significant parts of the analysis into 4 notebooks:

**Exploration.ipynb**

It contains the exploratory analysis performed, dataset encoding, and filtering. This was a crucial part of the process because, in addition to removing some features, it was observed that the original dataset, especially for one of the two labels under consideration, was quite unbalanced towards the negative class, containing those customers not identified as potential buyers. The latter are customers who were not identified as potential buyers of either type of product, and considering our primary goal is to sell, they could affect the performance of our classifiers. In the presentation, it's shown that, with almost equal accuracy, a higher recall is achieved by removing such customers in the particular case of multi-layer perceptron. Our choice was therefore to proceed only with those customers for which a purchase preference can be expressed.

**Classification.ipynb**

In this notebook, two classification tasks were carried out simultaneously and independently, one for the 'IncomeInvestment' label and the other for 'AccumulationInvestment'. For both, 9 different classifiers were trained (refer to the notebook for details) and then the 5 most promising ones were selected in terms of recall score, while also considering the accuracy score. We chose to give more weight to recall since our primary goal is to identify potential buyers belonging to the positive class, and we deemed it more critical to miss a potential buyer than to misidentify someone who is not.

With the 5 selected classifiers, we then obtained the true classification model through majority voting in the predictions of the individuals. The objective of this procedure is to obtain a more performant and robust ensemble model with lower variance than the individual classifiers, so that it can perform well even in the presence of noisy data.

We also collect the theory behind each classifier used in the file ‘TheoryRecap.pdf’.

**ClassificationStatistics.ipynb**

Once the global model for both classification tasks was obtained, it was tested on a portion of data not used for model training. In this notebook, statistics were derived regarding the samples classified as positive and negative in both cases. The goal here is to identify the key characteristics of the prototype customer who can potentially buy one or the other type of product; for the results, refer to the notebook or to the presentation.

**Recommendation.ipynb**

Finally, once the classification was done, we proceeded with a basic product recommendation, where customers classified as potential buyers in one or the other category are offered the respective product if it has an acceptable risk factor compared to the customer's risk propensity. Then, noticing that following this strategy left some potential buyers without a proposal due to their low risk propensity, the idea was to also propose an alternative, where they are suggested the product of the corresponding category with the lowest risk factor.