

You are provided with a list of clients. For each client, you are given with their height, salary, and whether this client made a purchase. Your task is to implement a set of classifiers, to predict if a future client will make a purchase.

For your prediction, you should train the following models:

- K-nn (based on Euclidian distance)
- Logistic regression
- Support Vector Classifier
  - Linear SVC
  - Polynomial SVC of degree  $m$  ( $m$  will be set in the code)
  - Gaussian SVC



### What to submit?

Python code implementing the task and the data generated for the scientific report (no CV is needed).

A scientific report (as a Word file):

- For the k-nn, evaluate the effect of  $k$  on the f1-score. For this evaluation, make 1,000 random train-test splits and provide the mean and STD f1-score for each  $k$  in the range 1 to 20 (included). **Specify what value of  $K$  will be chosen and why**
- F1-score for the Logistic regression and Linear SVC models. For this evaluation, make 1,000 random train-test splits and provide the mean and STD f1-score (for each model).
- For the polynomial SVC, evaluate the effect of the degree ( $m$ ) on the f1-score. For this evaluation, make 1,000 random train-test splits and provide the mean and STD f1-score for each  $m$  in the range 2 to 5 (included). **Specify what value of  $m$  will be chosen and why**
- For the gaussian SVC, evaluate the effect of  $C$  on the f1-score. For this evaluation, make 1,000 random train-test splits and provide the mean and STD f1-score for each  $c$  in the following set of values (0.2,0.5,1.2,1.8,3). **Specify what value of  $C$  will be chosen and why**

To make a fair analysis (and future comparison of the models), make sure to use the same 1,000 train-test splits to evaluate all models and their hyperparameters.