Let us imagine that we are a company, specializing in online retail sales, with brand name athletic footwear and accessories (e.g., Nike, Adidas, Zalando, etc). To better serve our visitors, e.g., suggest them products that they might be interested in buying, we track all users' activities across our websites. In particular, we record all visit sessions, and pay more attention to those that led to a purchase by storing their purchase time since last visit, as well as many other features. From the collected data, the Head of Product wanted to estimate the probability that a user will buy a specific product when visiting our websites. He then spoke to Mr. Google, who advised him to hire **a talented Master Student from ESCP school**, in order to transform his idea into reality. And now, you understand why we are sending you this exercise!

So, he's asked you to build a machine learning application, which can help us predict the probability of purchase for each line in our dataset. The dataset contains two files named train.csv (to help you train your model) and test.csv (see readme.txt for more details). Once your model is trained, you have to use the test.csv file to test your model, and send us the results as a csv file containing only two columns: id, prob. For example (see the prediction.csv file), your submission file should look like:
Id, Prob
5002, 0.5
5003, 0.8

Note that: you must also submit your code/solution. We combined both assignment 2 and assignment 3 into this test.

If you have any questions, do not hesitate to send me an email.

Anh-Phuong TA