Project number 1 (deadline : 07/02/2019)

We want to use a naive Bayes classifier (NBC) to predict the author of a mystery text.

— Download the files on the course webpage [1]. This zip files contains novels of Maupassant and Jules Verne, plus a mystery novel, whose authorship we want to assign to Maupassant or Verne (without reading the text . . .).

— Use your favorite search engine to learn how to handle text files and strings with python (see the footnote for a useful page [2]).

— Build a NBC classifier in order to predict the author of the mystery text (you can use any trick, expanding the training set by cutting the novels in pieces is authorized if you find it useful).

— Write a report explaining your approach and including the code you produced. The report has to be in pdf and sent by e-mail on the deadline (or before). You can work with other students but you have to give me one report per student.

---

1. `https://math.unice.fr/~rubentha/enseignement/ml-tp4-fichiers.zip`
2. `https://www.pythonforbeginners.com/files/reading-and-writing-files-in-python`