README

Our work is divided into:

* two notebooks, one for pre-processing and one for creating models
* a file “g6\_all\_functions.py” containing all the functions created and used in the notebooks
* a data folder containing the initial data file from the scrapping group, a data file transformed into tf-idf for each target variable, a data file transformed into word embedding for each target variable, a gloves file used for the word embedding , a post tagging file on 6000 lines and a table in the form of an image presented in the model notebook.
* A table “g6\_models\_results” showing all the results obtained for each target variable with all the algorithms tested and the parameters used.

It is not useful to launch the “g6\_pred\_preprocessing” notebook because all the output dataframes have been saved in csv in the data folder in order to be imported into the “g6\_work\_modeling” notebook to save time.

The “g6\_work\_modeling” notebook when it can be launched quickly enough, it is necessary to define the folder where the data is located in the “path”, and the one where the functions file is located.

However, all of the code has been launched and the results are already displayed and can be viewed.