

# Pa1 Report (b07801004)

## Environment:


Jupyter Notebook


## Langage:

Python3


## Execute:

1. Install nltk.(if needed)
2. Require “input.txt” and “NLTK\_stopWord.txt” before running the program.

 input.txt

 NLTK\_stopWord.txt

3. Run pa1.ipynb.
4. Output will be in “result.txt”.

 result.txt

## Discription:

1. Read “input.txt”, and split it by space.
2. Remove 16 kinds of punctuations.
3. Lowercase everything.
4. Stem using Porter’s algorithm(with The Natural Language Toolkit (NLTK) , a Python package).
5. Remove stopwords (with NLTK's list of english stopwords).
6. Write the result into txt file, “result.txt”.