

The submission contains the python source code and predicted file including others for the model described in:

*“FIT5149 Assessment 2: Sentiment Classification for Product Reviews”*

The directory contents of this distribution should as follows:

./

README - This documentation

group42\_ass2\_impl.ipynb - File containing the source code

This file runs the ULMFit model on the input data and predicts the test data producing the output file 'predict\_label.csv'

This is a python-based implementation. The environment used for building it was Python 3.6.4 and Google Colab (with GPU). The notebook can be compiled using using Jupyter Notebook 5.4.0 (64-bit) or higher as well.

[Not that: the code will take hours to build the model and predict with a decent GPU.]

=====

### Running Instruction

1. Run all the cells in 'group42\_ass2\_impl.ipynb' sequentially since input of next cell depends on the output of previous cell.
2. Please download the required libraries before compiling the code
3. It is a good idea to run the jupyter file in Google Colab rather than using any other platform for jupyter notebook or local machine.

=====

### Input files

The input data are as follows. These files should be in the same folder as the source code file:

./

labeled\_data.csv - File containing the labelled data (not submitted)

This file should contain 2 columns with names 'label','text' with first row specifying the column names

test\_data.csv - File containing the test data (not submitted)

This file should contain 2 columns 'test\_id', 'text' with first row specifying the column names

=====

Output file

The output file generated are:

./

predict\_label.csv - File containing the predicted values. (submitted as well)

This file contains the label prediction on the testing data. It contains 2 columns 'test\_id', 'label' with first row specifying the column names.

=====

The code can be run on Google colab or using local machine.

1. Local machine: The files ('group42\_ass2\_impl.ipynb', input files) should be in one folder and code will run (download the required packages beforehand).
2. Google colab:
  - a. The files should be uploaded to google drive in one folder
  - b. When running the code in colab, uncomment the first block of code and give a link to the google drive (already in the code).