# Project Title

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

# Directory contents

This package contains the Python code, the output file “predict\_label.csv” and the group assessment cover sheet.

The directory contents of this distribution are as follows:

|  |  |
| --- | --- |
| FIT5149 S2 2019 - A2 - Group 21.ipynb | Python code for Data Preparation & Feature Extraction, Model Classifier and Model Evaluation |
| README | This documentation |
| predict\_label.csv | The label prediction on the test dataset |
| assignment-cover-group.pdf | The signed group assignment cover sheet |

# Running Instruction

**Jupyter Notebook**

* The user would require to upload the “FIT5149 S2 2019 - A2 - Group 21.ipynb” file on the root folder.
* Once the file is opened in Jupyter Notebook then user can click on “Kernel” from the menu option, then click “Restart & Run All” to run all cells of the file sequentially. It will import all required libraries and install any libraries that are required to process the code.

**Google Colaboratory (Recommended)**

* The user would require to upload the “FIT5149 S2 2019 - A2 - Group 21.ipynb” file on Google Colaboratory. This can be done via the File menu then upload notebook.
* Similarly, “FIT5149 S2 2019 - A2 - Group 21.ipynb” can first be uploaded on Google drive then open it via Google Colaboratory.
* Then from the menu option the user can click on “Runtime” and then click “Run all” or press “Ctrl+F9” on keyboard to run all cells of the file sequentially. It will import all required libraries and install any libraries that are required to process the code.

# Input data files

The “FIT5149 S2 2019 - A2 - Group 21.ipynb” requires three input data files “labelled\_data.csv”, “unlabelled\_data.csv” and “test\_data.csv”.

The “labelled\_data.csv” and “unlabelled\_data.csv” will be used for feature extraction and model building while the “test\_data.csv” will be used to make label prediction by the model.

# Output file

Once the “FIT5149 S2 2019 - A2 - Group 21.ipynb” file is finished processing it will produce “predict\_label.csv” file which will contain the label prediction on the test dataset.