Manual for Bug Report Classification program

The GitHub directory contains 3 python files (br_baseline_NB.py, br_classification.py, and br_classification_gridsearch.py), the datasets directory and a requirements.txt file.

Here is what each of these files do:

br_baseline_NB.py - This is the baseline tool provided in lab1. This is here for comparison between the baseline tool and my solution.

br_classification.py - This is the main training file for the LSTM approach to bug report classification.

Br_classification_gridsearch.py - This is the file used for performing the gridsearch on the network whilst training to find the optimal hyperparameters.

How to use the tool:

- First, create a virtual environment in the parent directory using the command python -m venv venv
- Connect to the python environment:
    - Windows - ./venv/Scripts/Activate.ps1
    - Mac & Linux - source venv/bin/activate
- Install requirements - pip install -r requirements.txt
- To run the gridsearch file:
    - Select the dataset you would like to run on, change line 78 in the br_classification_gridsearch.py file to the desired dataset from the datasets directory.
    - Run the br_classification_girdsearch.py file.
    - The pre-trained GloVE embeddings and stopword files will be automatically downloaded.
    - Once training has finished, the results containing the grid search results will be wrote to the grid_search_results.txt file.
    - To experiment with any other values in the gridsearch such as widening the search space, please change lines 173-177 which contains the grid search parameter grid.
- To run the training file:
    - Select the dataset you would like to run on, change line 80 in the br_classification_gridsearch.py file to the desired dataset from the datasets directory.
    - Run the br_classification.py file.

    - If not downloaded already, this will automatically download the GloVE embeddings and stopword packages.
    - Once training has finished, the results will be printed to the console.
- For running the baseline tool, this file is a copy from the lab1 solution, so execution is exactly the same.