Read Me

Source Code Description:

There is no dataset file included as the when the code is executed the dataset will be downloaded on the host machine from the official dataset link.

The source code folder has 6 .py files which only contain the python code and 6 .ipynb files which contain the python code and output for the classifiers. All files are independent and can be executed individually.

Naming convention for the files is as follows:

<<**WordEmbeddingTechnique**>>_classes_<<**Number of classes**>>

For e.g. the code for the file containing the output for CountVector as the word embedding technique and having 3 ratings to classify has been named as:

countvector_classes_**3**.ipynb and **countvector**_classes_**3**.py

All files contain implementation and output for all 3 classifiers used

Note The dataset contains over 9 million rows we had memory issues and some sections took multiple hours to execute so at regular intervals progress is saved to google drive as this project was executed using google collab. As a result certain directory path need to be updated. Example of path that need to be updated is shown below:

```
from google.colab import drive
drive.mount('/content/drive')
datasetPath = "/content/drive/My Drive/Natural Language Processing/Project/Cleaned Dataset/"
#Saving to CSV File
```

Execution Instructions:

Follow the below instructions to run the code:

1) Install python and pip

2) Go to the source code folder run the command " pip install requirements.txt " in order to install all the dependencies

3) Open a py or ipynb file change all the directory paths as required

4) Execute the code

Note: The dataset contains over 9 million rows we had memory issues and some sections took multiple hours to execute. So it was not possible to execute the entire code at once. As a result not all blocks have outputs in the .ipynb. But the main classifier blocks all have their output for all the files