Our code base has been written on Sublime text for simplicity.

To regenerate the machine learning model:

The src folder has the following contents:

1. **dataset.csv** - this contains our data. It has a number of essays and corresponding scores for each essay.
2. **project.py** - this python file uses the Word2Vec model for training.
3. **tf-idf.py** - this python file uses the tf-idf scores for training.
4. **tf.pickle** - a pickle file which stores the Vector list. Vector list is list of vectors (one vector for each essay) created using tf-idf. In our program (**tf-idf.py**) we just read from the pickle file to save time
5. **wvl.pickle** - stores the Word Vector List, which is list of ‘list of words for each essay’.
6. **word2vec\_model.pickle** - stores the Word2Vec model

Instructions to run:

All of the code has been written using Python version 3.6.5.

Important python libraries required,

1. nltk
2. gensim
3. keras
4. pickle
5. sklearn
6. numpy, pandas

The following terminal commands can be run to execute our code.

$ python3 project.py

$ python3 tf-idf.py

Dependencies for the server.py(Web tech file)

1. **testdata.csv-** this is the single essay test input data

2. **my\_model\_w2v.h5**- The LSTM word2Vec model is stored here

Terminal command to execute server.py

$ python3 server.py