Software Documentation

1) Function of the code

This program I wrote calculates the EOWC for a user inputted review and every Yelp review from the training data. It then calculates the dot product for each rating (i.e. 1, 2, 3, 4, 5) and returns the rating with the highest dot product. The program can be used to estimate the rating for some text. For example, you could try to input “Very tasty”, and the program would output the most likely rating associated with it.

2) Documentation of how the software is implemented with sufficient detail so that others can have a basic understanding of your code for future extension or any further improvement

I wrote two programs. The first one trains the model, and the second one is an interactive application that is used as a predictor. Model_trainer.py is a trainer, where it takes in a body of text and a rating. It then tokenizes the words and logs it into data.csv. To keep the prediction fair, it is normalized by dividing the total count of words by the number of reviews with that given rating. This allows the model to have an estimated word count per review for a given rating.

Rating predictor is an interactive python application that can predict a rating given a body of text. It calculates the dot product with all the reviews in the training data and returns the rating with the highest corresponding dot product.

Future extension/improvements: Currently, words that are synonyms are not considered the same. As an extension, I could use a cloud-based free-text search solution to calculate similarities between words before calculating the dot product. This would give a more accurate prediction. For example, currently the dot product between “Tasty” and “Delicious” is zero, but after comparing these words, the cloud-based free-text search solution should give a non-zero score.

3) Documentation of the usage of the software including either documentation of usages of APIs or detailed instructions on how to install and run a software, whichever is applicable.

Steps:

1. Download folder from https://github.com/timothycheung1/CourseProject
2. If you want to train your own data:
    a. Run model_trainer.py and:
    b. Insert reviews along with corresponding rating
3. Run rating_predictor.py:
    a. Type review
    b. See what the model predicts

4) Brief description of contribution of each team member in case of a multi-person team

N/A, I worked by myself.