Shikun Lyu

000879956

1. The first step is to read the csv file of three datasets.
2. Delete the first column of the test\_set.csv.
3. Get the sentences, words element and put into dictionary.
4. Vector the dictionary for 2031 length list.
5. KNN classification.
6. Do Accuracy based on the valid dataset.
7. The best K is 17 with almost 39% accuracy rate. Test K values from 1 to 20 for odd.
8. Use the KNN to label the test\_set.csv file.
9. Output the csv file with the label inserted.