CMPE 239 HW02

Basics

- Name: Shiyu Mou
- ID: 011551358
- Rank & Accuracy: Top7 and 0.7819

Methodology

- Combined the tranning set and the test set together to build the dictionary
- Filtered short words that shorter than 4
- Built a sparse matrix
- Used idf to decrease the importance of popular words
- Normalized the matrix
- Computed the cosine similarity between Ys and Xs
- `sims = Y.dot(X.T)`
- Sort the matrix using np.argsort, pretty much faster than sorting a dictionary with sorted()
- Rank the top 100 neighbours and vote for the most labels.
- Give every test data a label with the most voted result.
- print to a txt file.