Aspect Extraction Assignment Report

Data:

Dataset is downloaded from Kaggle.com. It contains reviews, rating, review_date,review_text,author,author_url,location,rating,food,value,service,visited_on,location etc.

Data Cleaning:

It includes removal of stop words, dropping the null value columns, removing float, integer value from *review_text* column only string is allowed.

Tokenizing:

Nltk library is used to get nouns and adjective of English language and then each word of review is labelled with the proper tag.

Noun => tag is "NN"

Noun Phrase => tag is "NNP"

Aspect Extracting:

Those words which are labelled by NN and NNP are used as aspects. So for every review aspects are extracted as *new_aspect_list* and added to list of all_aspects if extracted length of aspect list is greater than 0.

(https://www.kaggle.com/PromptCloudHQ/londonbased-restaurants-reviews-on-tripadvisor/data)