

Assumptions

1. Labelled testing data file (SubtaskA_Trial_Test_Labeled.csv) has been used for calculating accuracy
2. For Word2Vec based classification, the process is taking a lot of time. Hence only the top 100 sentences have been predicted to sentences or not

Learnings

1. We need a lot of training data for the successful application of word2vec model of data. This is because Word2Vec model is able to create a contextual meaning from the data only when it has got ample amount of data.
2. Word2Vec combined with wmdistance takes a lot of time hence its application is not feasible all the time. Eg. this dataset
3. Naive Bayes is a simple and effective technique for text classification. It's application is simple and it gives good result for our data
4. We can mine suggestions in the above simple and effective ways which can be beneficial when predicting hotel review data