Topic Modeling

- The attached zip has 2964 files with text data. They are documents belonging to 3 topics – Sports, Technology and Religion
- Do an unsupervised clustering of documents into 3 clusters
- The actual topic of each file is provided as reference in target.csv
- Check whether the documents of same topic have been grouped together. Try with different feature extraction options (e.g: tfidf, removing stop words, stemming etc.) and see whether performance improves
- Now that you know the actual class label, try doing Supervised Classification
    - Divide the documents into training and test
    - Use different modeling algorithms to identify the topic
    - Test the algorithms using test data