Latent semantic analysis

First, we have taken 100 news articles from the Kaggle news data set.

We cleaned the data set to remove redundant words.

Then we converted the news article to term-doc matrix.

Then we used SVD (singular value decomposition) to reduce the features and get reduced vector.

Then we find the cosine similarity among all documents and prepare doc-doc matrix.

Now all the document with high similarity are kept in same group.