

Tools Used and All performed steps:

1-for preprocessing I used pandas for reading csv file , then I concatenated ("Abstract") and ("Title") to one column , then I removed any row its length less than 50, because I saw the there is not many rows its ("Abstract") column has no meaningful value.

2-Then I did any NLP design flow(Stemming , Lemmatization, TF-idf).

3-Then I used gensim.corpora.Dictionary to convert processed data frame to this Dictionary to use it when creating the LDA Model.

4-Then I made a function called filter_extremes to delete the words that occur 5 times or below and not appearing in more than 10 percent.

5-Then converting the Dictionary to a Bag of Words corpus.

6- Then I used genism.models.LdaMulticore to train the model and give it the number of topics equal to 7 because I wrote another function that made a grid search over the number of topics and I found that when number of topics is 7 made a high coherence score.

7-Then I used CoherenceModel to evaluate the model.

8-finally I used wordcloud to plot the weights of each words per topic.

How you can further improve the results:

1-we can make a grid search over hyperparameters alpha and beta .

2-hyperparameter tunning .