LDA:

* Different number of topics, compare perplexity, use the same dataset (split into training and test)
  + 10,20,30,40,50,70,100,150 topics
  + Calculate perplexity (check it)
  + Save the result to pickle: takes time to run LDA
* Different preprocessing, using the best number of topics, compare NPMI, (other metrics: silhouette coefficient, calinski-harabaz??)
  + Implemented NPMI
  + without Stop words, without stem
  + without Stop words, without stem
  + without stem, With Stop words,
  + With stop words, with stem
  + Product \_ id : many reviews - one topic

LDA running:

Xiaomeng: 100,150

Bagus: 50,70

Ayush: 10,20

Anupam: 30, 40