

Chen Eliyahou 312490675

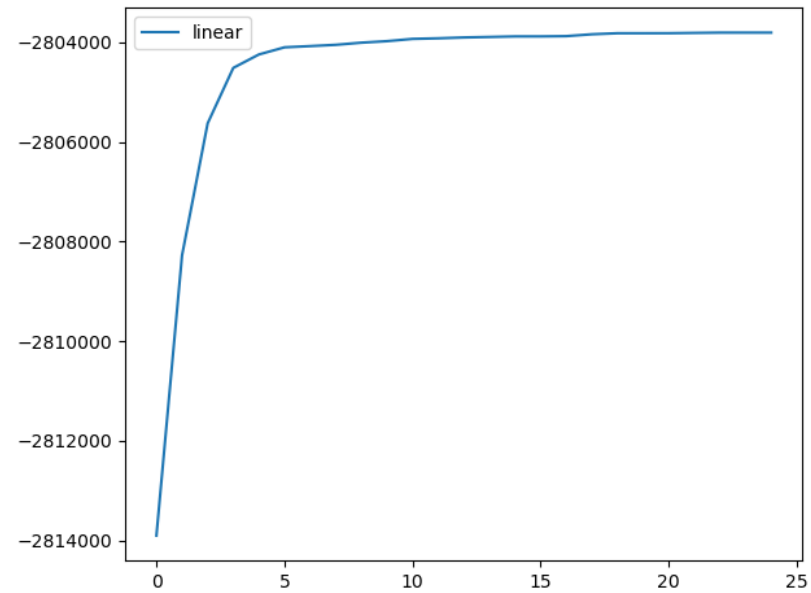
Noam Simon 208388850

Report:

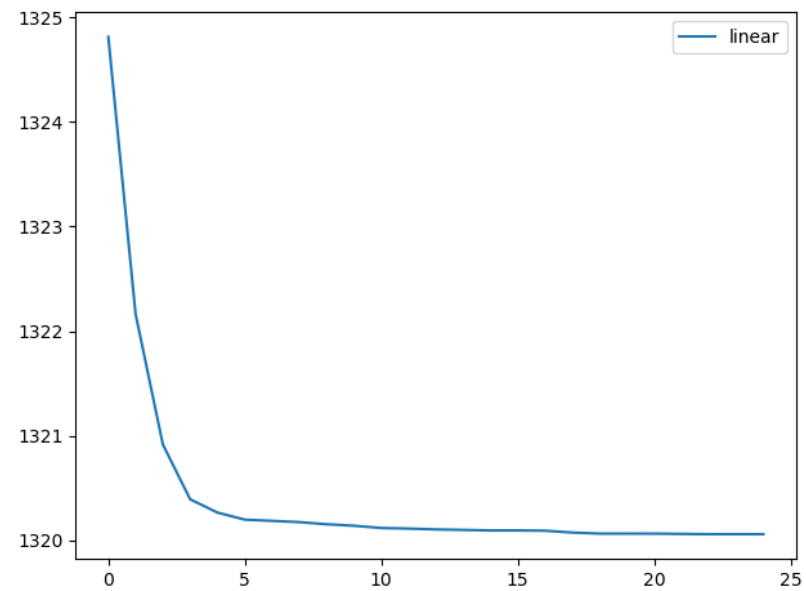
1. The threshold was a different likelihood value of 0.01 between iterations.

2.

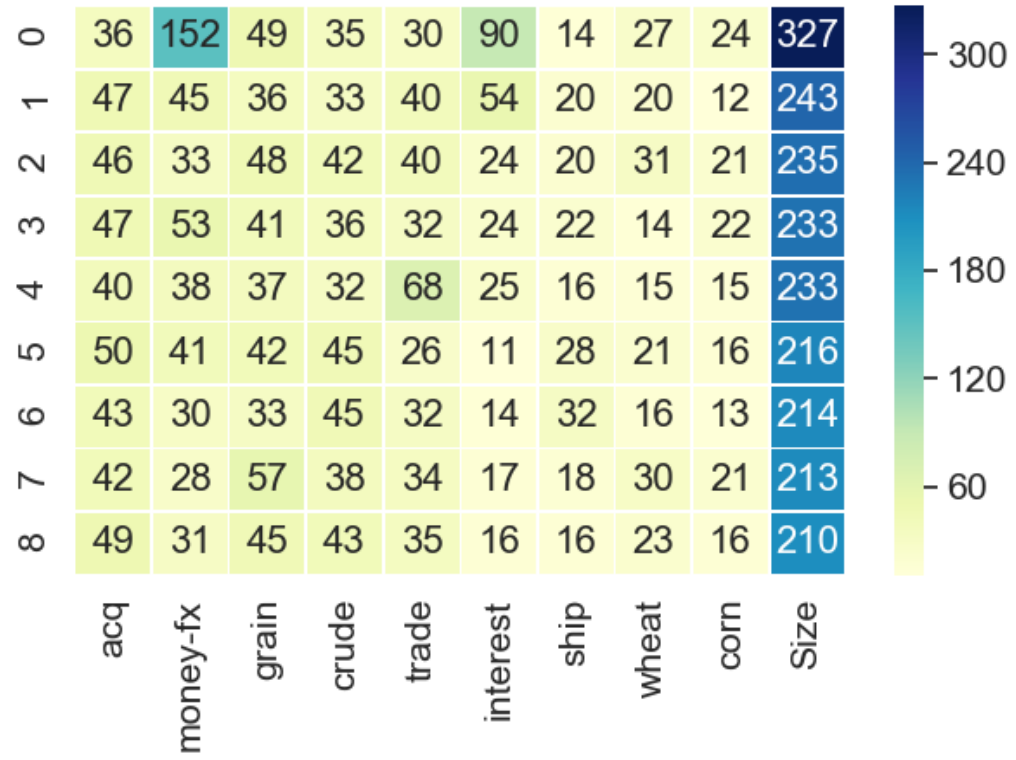
a. Likelihood per iteration



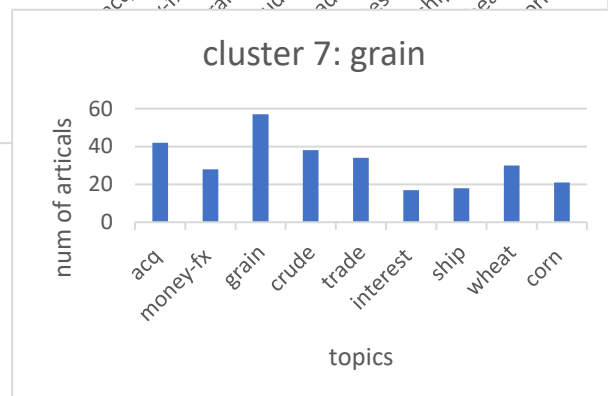
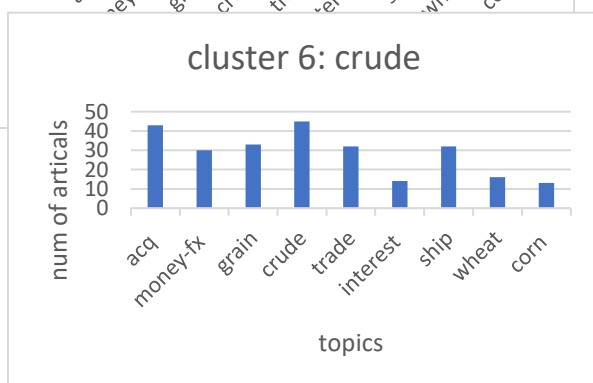
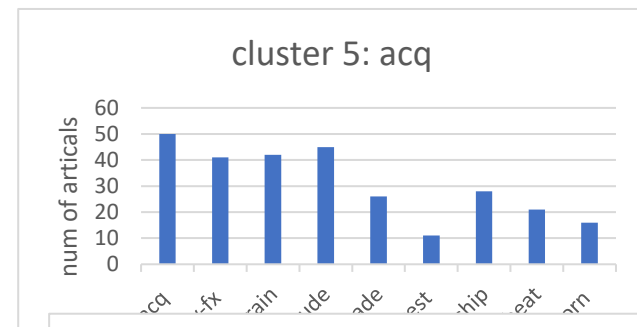
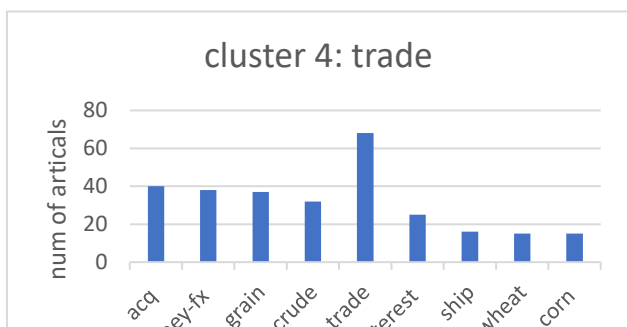
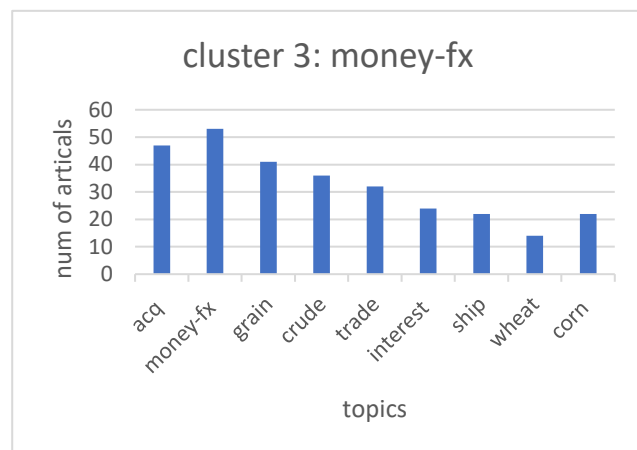
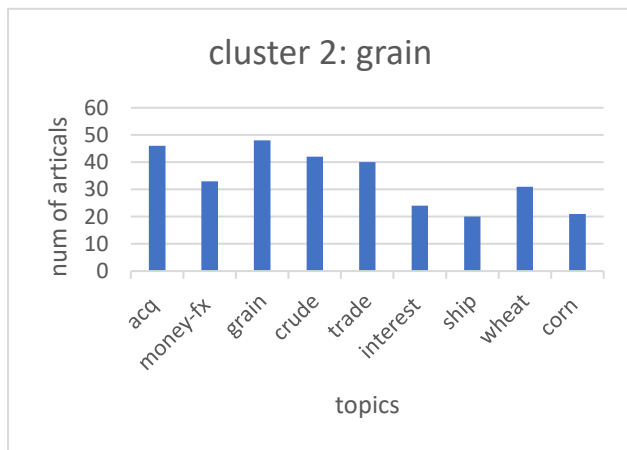
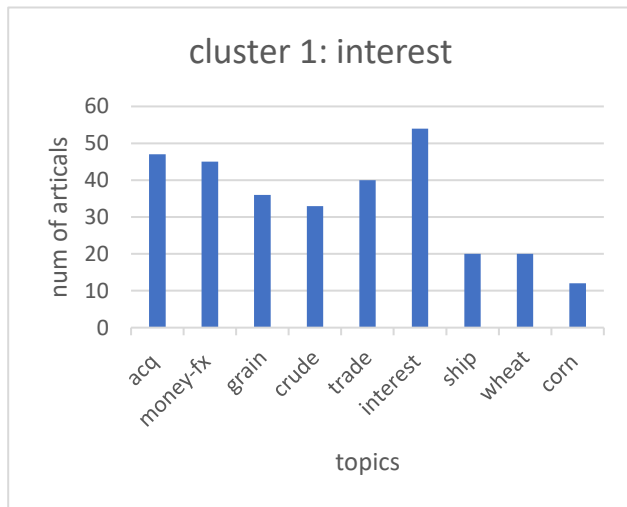
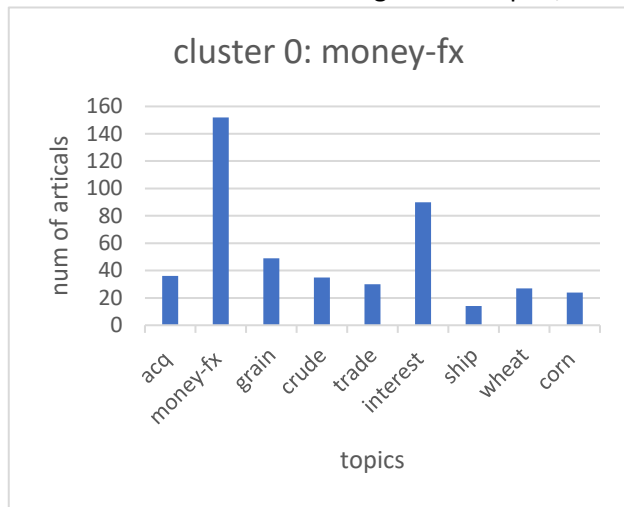
b. Perplexity pre iteration:

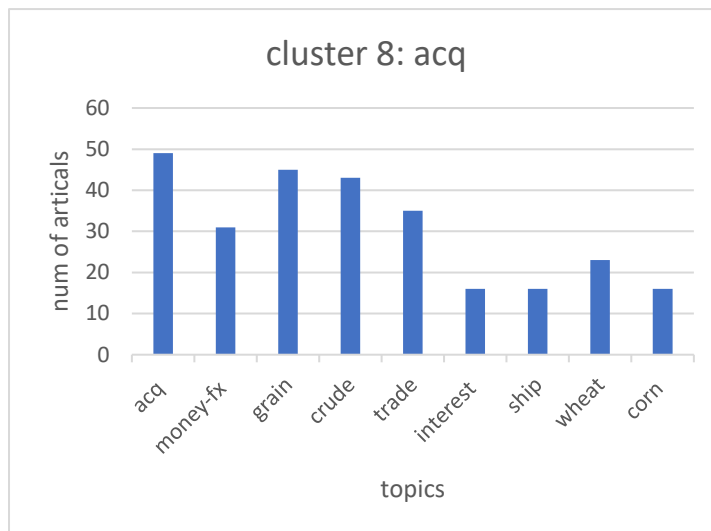


3. confusion matrix:



4. 9 histograms of topics, one for each cluster:





5. Acc: 0.2711864406779661

6.

- a. $K = 10$
- b. vocabulary size after filtering - 6,800
- c. $\lambda: 0.001$