Applied probability models for CS: Exercise 3

Daniel Bazar 314708181

Lior Krengel 315850594

## Part 1

We decided the threshold for stopping criteria in the EM algorithm to be **10.** In scale of the large numbers of the log-likelihood, 10 is a very small number. We tested many thresholds and there weren’t big differences.

## Part 2

Log-likelihood and mean perplexity plots until convergence (28 iterations):

A graph with a line drawn on it

Description automatically generated A graph with a line

Description automatically generated

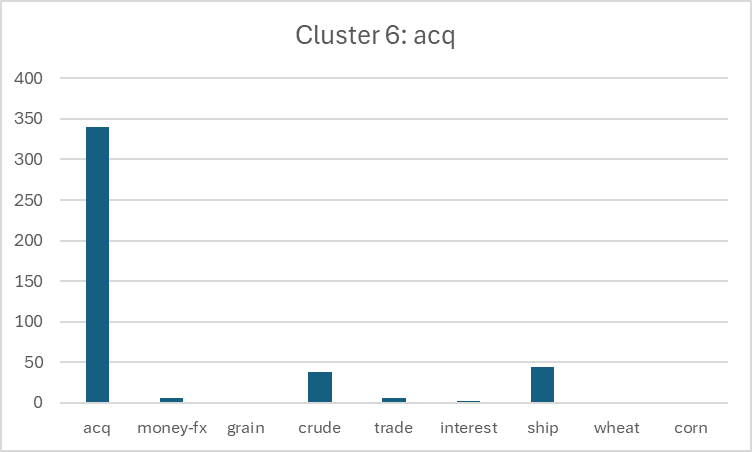
## Part 3

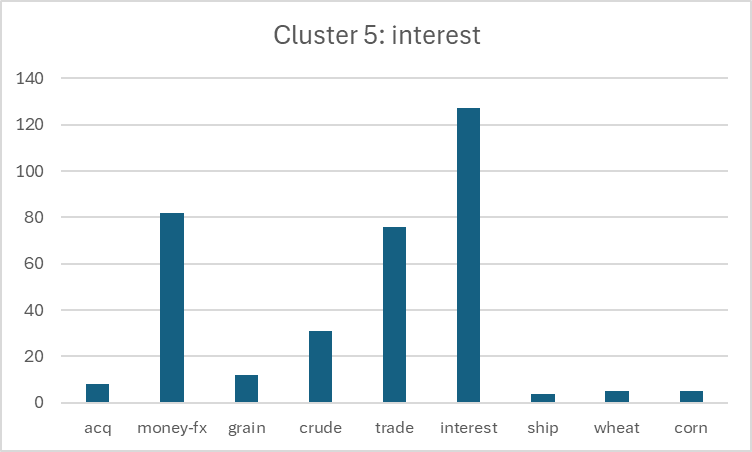
Confusion matrix:

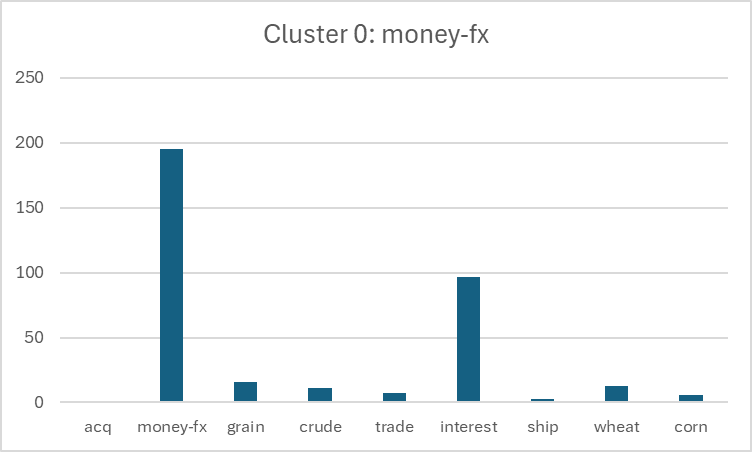
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | acq | money-fx | grain | crude | trade | interest | ship | wheat | corn | size |
| 6 | 340 | 6 | 1 | 38 | 6 | 2 | 44 | 1 | 0 | 438 |
| 5 | 8 | 82 | 12 | 31 | 76 | 127 | 4 | 5 | 5 | 350 |
| 0 | 0 | 195 | 16 | 11 | 7 | 97 | 3 | 13 | 6 | 348 |
| 7 | 0 | 9 | 144 | 30 | 13 | 2 | 12 | 79 | 56 | 345 |
| 2 | 7 | 7 | 92 | 43 | 19 | 12 | 24 | 60 | 33 | 297 |
| 8 | 8 | 47 | 15 | 14 | 183 | 7 | 4 | 3 | 7 | 288 |
| 3 | 2 | 71 | 63 | 21 | 19 | 12 | 22 | 22 | 40 | 272 |
| 1 | 13 | 22 | 35 | 34 | 6 | 8 | 64 | 11 | 11 | 204 |
| 4 | 22 | 12 | 10 | 127 | 8 | 8 | 9 | 3 | 2 | 201 |

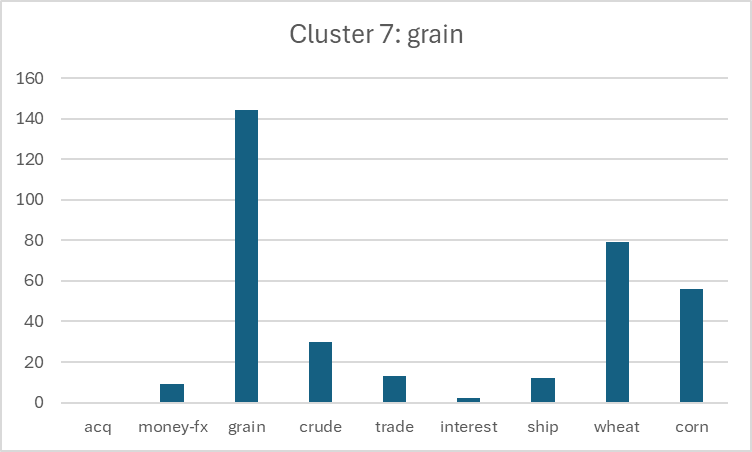
## Part 4

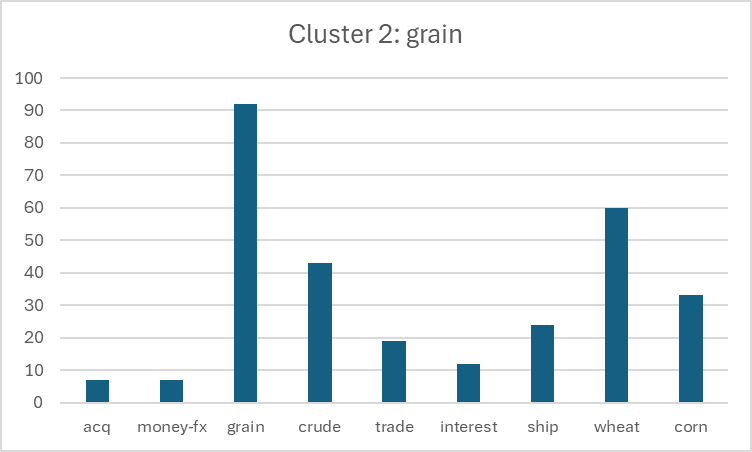
Topics histograms:

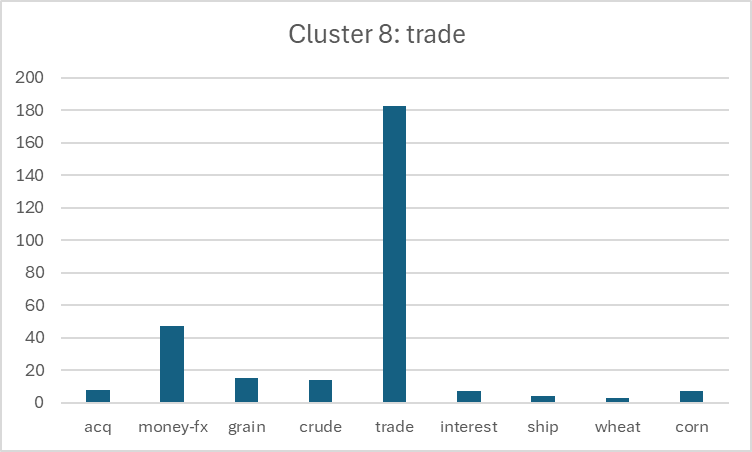


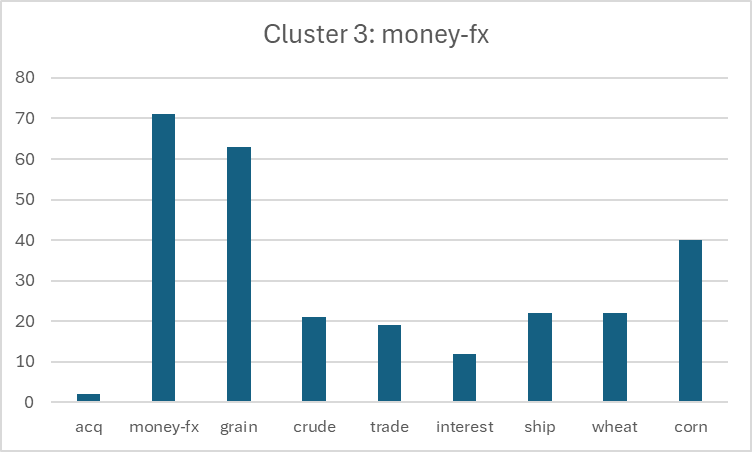


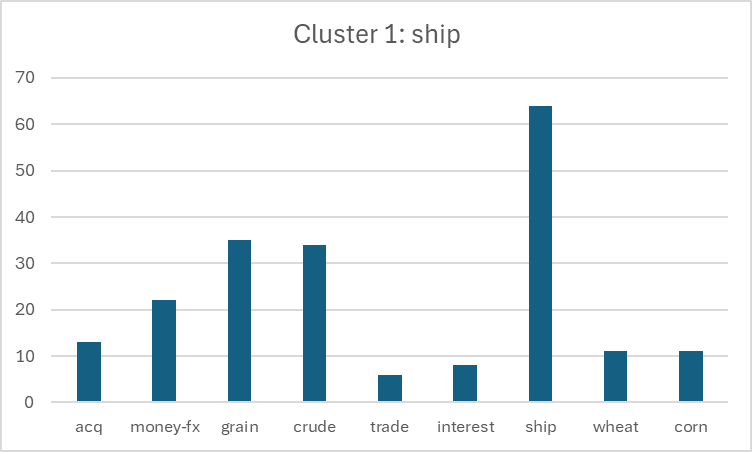


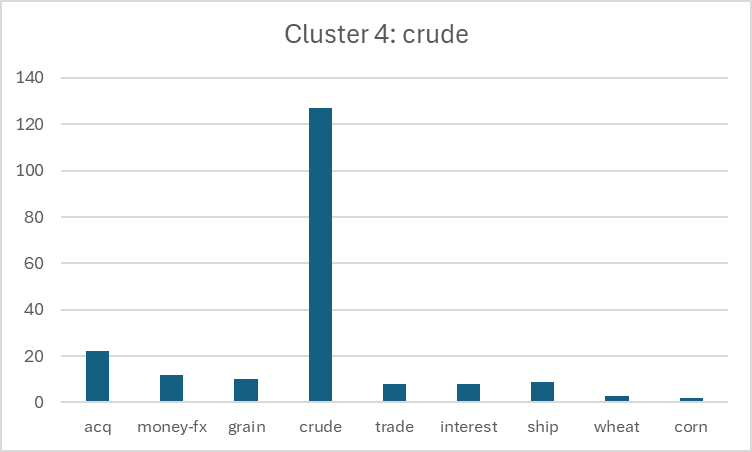












## Part 5

Model accuracy: **63.23%**

## Part 6

|  |  |
| --- | --- |
| K | 10 |
| vocabulary size after filtering | 6800 |
| λ | 0.06 |