



# Computational social media

**Lecturer :**

Gatica-Perez Daniel

Assignment #3

Negin Safaei

## Report on assignment #3:

### 2. Preprocessing:

- a. Converting text into lowercase.
- b. Removing non-alphabet characters.
- c. Breaking sentences into words.
- d. Removing short word (length less than 3).
- e. converting a word to its base form.
- f. Removing common English words.

### 3. Implementing LDA (n\_topics=20, n\_iter=1500, random\_state=1)

- a. Removing words that occur in less than 10 and more than 90% of the documents.
- b. Building a dictionary.  
Length of dictionary: 4404
- c. computing the frequency of each word.
- d. Applying the LDA model.

```
INFO:lda:n_documents: 50000
INFO:lda:vocab_size: 4404
INFO:lda:n_words: 415699
INFO:lda:n_topics: 20
INFO:lda:n_iter: 1500
WARNING:lda:all zero row in document-term matrix found
INFO:lda:<0> log likelihood: -4846181
INFO:lda:<10> log likelihood: -2992754
INFO:lda:<20> log likelihood: -2850711
INFO:lda:<30> log likelihood: -2821051
INFO:lda:<40> log likelihood: -2810245
INFO:lda:<50> log likelihood: -2803025
INFO:lda:<60> log likelihood: -2799259
INFO:lda:<70> log likelihood: -2796197
INFO:lda:<80> log likelihood: -2793214
INFO:lda:<90> log likelihood: -2791447
INFO:lda:<100> log likelihood: -2788975
INFO:lda:<110> log likelihood: -2789031
INFO:lda:<120> log likelihood: -2787932
INFO:lda:<130> log likelihood: -2787089
INFO:lda:<140> log likelihood: -2785746
INFO:lda:<150> log likelihood: -2785018
INFO:lda:<160> log likelihood: -2783960
INFO:lda:<170> log likelihood: -2780791
INFO:lda:<180> log likelihood: -2781155
INFO:lda:<190> log likelihood: -2779919
INFO:lda:<200> log likelihood: -2778245
INFO:lda:<210> log likelihood: -2777076
INFO:lda:<220> log likelihood: -2776878
INFO:lda:<230> log likelihood: -2775837
INFO:lda:<240> log likelihood: -2775040
```

INFO:lda:<250> log likelihood: -2774341  
INFO:lda:<260> log likelihood: -2773055  
INFO:lda:<270> log likelihood: -2772682  
INFO:lda:<280> log likelihood: -2772505  
INFO:lda:<290> log likelihood: -2772634  
INFO:lda:<300> log likelihood: -2772208  
INFO:lda:<310> log likelihood: -2772722  
INFO:lda:<320> log likelihood: -2771580  
INFO:lda:<330> log likelihood: -2772511  
INFO:lda:<340> log likelihood: -2771748  
INFO:lda:<350> log likelihood: -2771612  
INFO:lda:<360> log likelihood: -2770349  
INFO:lda:<370> log likelihood: -2770179  
INFO:lda:<380> log likelihood: -2771291  
INFO:lda:<390> log likelihood: -2770463  
INFO:lda:<400> log likelihood: -2770130  
INFO:lda:<410> log likelihood: -2770510  
INFO:lda:<420> log likelihood: -2770564  
INFO:lda:<430> log likelihood: -2770019  
INFO:lda:<440> log likelihood: -2769412  
INFO:lda:<450> log likelihood: -2770192  
INFO:lda:<460> log likelihood: -2769678  
INFO:lda:<470> log likelihood: -2769186  
INFO:lda:<480> log likelihood: -2769550  
INFO:lda:<490> log likelihood: -2768999  
INFO:lda:<500> log likelihood: -2768932  
INFO:lda:<510> log likelihood: -2768242  
INFO:lda:<520> log likelihood: -2769430  
INFO:lda:<530> log likelihood: -2769419  
INFO:lda:<540> log likelihood: -2768258  
INFO:lda:<550> log likelihood: -2768582  
INFO:lda:<560> log likelihood: -2768477  
INFO:lda:<570> log likelihood: -2768160  
INFO:lda:<580> log likelihood: -2767284  
INFO:lda:<590> log likelihood: -2767918  
INFO:lda:<600> log likelihood: -2768739  
INFO:lda:<610> log likelihood: -2766785  
INFO:lda:<620> log likelihood: -2767451  
INFO:lda:<630> log likelihood: -2766649  
INFO:lda:<640> log likelihood: -2766918  
INFO:lda:<650> log likelihood: -2767561  
INFO:lda:<660> log likelihood: -2767276  
INFO:lda:<670> log likelihood: -2765282  
INFO:lda:<680> log likelihood: -2765530  
INFO:lda:<690> log likelihood: -2766792  
INFO:lda:<700> log likelihood: -2766848  
INFO:lda:<710> log likelihood: -2766342  
INFO:lda:<720> log likelihood: -2766472  
INFO:lda:<730> log likelihood: -2767072  
INFO:lda:<740> log likelihood: -2767209  
INFO:lda:<750> log likelihood: -2766254  
INFO:lda:<760> log likelihood: -2766473  
INFO:lda:<770> log likelihood: -2765731  
INFO:lda:<780> log likelihood: -2766471  
INFO:lda:<790> log likelihood: -2765759  
INFO:lda:<800> log likelihood: -2765953  
INFO:lda:<810> log likelihood: -2766975  
INFO:lda:<820> log likelihood: -2764944  
INFO:lda:<830> log likelihood: -2766614  
INFO:lda:<840> log likelihood: -2765762  
INFO:lda:<850> log likelihood: -2766041

INFO:lda:<860> log likelihood: -2765285  
INFO:lda:<870> log likelihood: -2765558  
INFO:lda:<880> log likelihood: -2765421  
INFO:lda:<890> log likelihood: -2765897  
INFO:lda:<900> log likelihood: -2765215  
INFO:lda:<910> log likelihood: -2765485  
INFO:lda:<920> log likelihood: -2765939  
INFO:lda:<930> log likelihood: -2764952  
INFO:lda:<940> log likelihood: -2765872  
INFO:lda:<950> log likelihood: -2765567  
INFO:lda:<960> log likelihood: -2765041  
INFO:lda:<970> log likelihood: -2764921  
INFO:lda:<980> log likelihood: -2765224  
INFO:lda:<990> log likelihood: -2764024  
INFO:lda:<1000> log likelihood: -2764489  
INFO:lda:<1010> log likelihood: -2764608  
INFO:lda:<1020> log likelihood: -2765186  
INFO:lda:<1030> log likelihood: -2764985  
INFO:lda:<1040> log likelihood: -2764889  
INFO:lda:<1050> log likelihood: -2764682  
INFO:lda:<1060> log likelihood: -2764854  
INFO:lda:<1070> log likelihood: -2764391  
INFO:lda:<1080> log likelihood: -2763910  
INFO:lda:<1090> log likelihood: -2764972  
INFO:lda:<1100> log likelihood: -2764014  
INFO:lda:<1110> log likelihood: -2764151  
INFO:lda:<1120> log likelihood: -2765329  
INFO:lda:<1130> log likelihood: -2764375  
INFO:lda:<1140> log likelihood: -2765394  
INFO:lda:<1150> log likelihood: -2764758  
INFO:lda:<1160> log likelihood: -2764798  
INFO:lda:<1170> log likelihood: -2764125  
INFO:lda:<1180> log likelihood: -2764452  
INFO:lda:<1190> log likelihood: -2764352  
INFO:lda:<1200> log likelihood: -2764488  
INFO:lda:<1210> log likelihood: -2763715  
INFO:lda:<1220> log likelihood: -2764120  
INFO:lda:<1230> log likelihood: -2764315  
INFO:lda:<1240> log likelihood: -2764178  
INFO:lda:<1250> log likelihood: -2764621  
INFO:lda:<1260> log likelihood: -2763890  
INFO:lda:<1270> log likelihood: -2764769  
INFO:lda:<1280> log likelihood: -2765531  
INFO:lda:<1290> log likelihood: -2764828  
INFO:lda:<1300> log likelihood: -2764580  
INFO:lda:<1310> log likelihood: -2765263  
INFO:lda:<1320> log likelihood: -2765021  
INFO:lda:<1330> log likelihood: -2764275  
INFO:lda:<1340> log likelihood: -2764087  
INFO:lda:<1350> log likelihood: -2765133  
INFO:lda:<1360> log likelihood: -2765341  
INFO:lda:<1370> log likelihood: -2765002  
INFO:lda:<1380> log likelihood: -2764678  
INFO:lda:<1390> log likelihood: -2764248  
INFO:lda:<1400> log likelihood: -2764876  
INFO:lda:<1410> log likelihood: -2764984  
INFO:lda:<1420> log likelihood: -2763933  
INFO:lda:<1430> log likelihood: -2764210  
INFO:lda:<1440> log likelihood: -2763901  
INFO:lda:<1450> log likelihood: -2763867  
INFO:lda:<1460> log likelihood: -2763668

```
INFO:lda:<1470> log likelihood: -2764928
INFO:lda:<1480> log likelihood: -2764665
INFO:lda:<1490> log likelihood: -2764265
INFO:lda:<1499> log likelihood: -2763965
<lda.lda.LDA at 0x125888110>
```

e. Examine the topics, with your own proposal for the “name” of each topic that is meaningful to you. Display your results in the notebook.

Topic 0: coronavirus news

Topic 0: death breaking spike past toll 24hrs piersmorgan number total fatality recorded horribly

Topic 1: iran

Topic 1: http covid19 briefing medium drtedros live conference iran press watch covid pandemic

Topic 2: trump lecture

Topic 2: trump realdonaldtrump china country warned death navarro people threat peter would tomfitton

Topic 3: boris johnson

Topic 3: http boris johnson minister care prime know intensive borisjohnson emergency state japan

Topic 4: news world

Topic 4: death south korea http rate trump world people first state crisis davidcorndc

Topic 5: quarantine

Topic 5: http stay people medical nurse ajenglish staff home tackle continues indoors urging

Topic 6: iran drastic situation

Topic 6: people crisis current continues suffering accelerate drastically occupant teampelosi iran maga regime

Topic 7: number of covid19 cases

Topic 7: case death http total confirmed covid19 number covid report positive reported update

Topic 8: stayhome

Topic 8: http covid19 covid china mask corona virus pandemic read covid2019 head stayhome

Topic 9: emergency

Topic 9: pandemic play point feeling realcandaceo watching arrived exact called patient quicktake help

Topic 10: italy

Topic 10: http lockdown covid start behind national pandemic today restriction italy social first

Topic 11: covid19 test

Topic 11: http testing state help covid19 tested patient working test combat positive covid

Topic 12: china spread covid19

Topic 12: http could crisis report spread china source lockdown november early intelligence warned

Topic 13: business in covid19 situation

Topic 13: http pandemic help covid19 spread need home business time outbreak health worker

Topic 14: USA shutdown

Topic 14: http latest trump advice daily thanks rule getting public president shutdown medical

Topic 15: lockdown world

Topic 15: people today help test million lockdown china contain none vietnam country cope

Topic 16: people

Topic 16: news worker england died patient hour health fake make home america grocery

Topic 17: quarantine life

Topic 17: http covid19 time lockdown week life lockdownextension whole imagine quarantinelif kordy90 keep

Topic 18: news

Topic 18: http every daily press world plan briefing shown brilliant fight utterly kthopkins

Topic 19: current situation

Topic 19: http covid19 lockdown pandemic update world already today place worldwide still confirmed

5. Discuss the results. Were any topics surprising? Use any of the material presented in class to support your discussion

Some of the topics seems irrelevant to each other and I could not find a name for them easily sometimes by considering more words for each topic finding a name becomes easier but sometime makes it worse and more difficult. According to our discussion in the class because our topic claim was just one topic (corona) every single topics are relevant to corona.