# Semi-supervised classification model: newsmap and keyATM

#### Miras Tolepbergen

2023-10-27

```
rm(list=ls(all=TRUE))
getwd()
## [1] "C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1"
setwd("C:/Users/Miras/Desktop/u m/1st/big data analytics/Labs/projects")
getwd()
## [1] "C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1"
library(manifestoR)
## Loading required package: NLP
## Loading required package: tm
## When publishing work using the Manifesto Corpus, please make sure to cite
it correctly and to give the identification number of the corpus version used
for your analysis.
## You can print citation and version information with the function
mp_cite().
##
## Note that some of the scaling/analysis algorithms provided with this
package were conceptually developed by authors referenced in the respective
function documentation. Please also reference them when using these
algorithms.
library(quanteda)
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "replValueSp"; definition not updated
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "xMatrix"; definition not updated
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "mMatrix"; definition not updated
## Package version: 3.3.1
## Unicode version: 13.0
## ICU version: 69.1
```

```
## Parallel computing: 4 of 4 threads used.
## See https://quanteda.io for tutorials and examples.
## Attaching package: 'quanteda'
## The following object is masked from 'package:tm':
##
##
       stopwords
## The following objects are masked from 'package:NLP':
##
       meta, meta<-
library(dplyr)
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
       filter, lag
##
## The following objects are masked from 'package:base':
##
       intersect, setdiff, setequal, union
##
# specify your key to get access to the CMP dataset via API
mp_setapikey(key.file = NULL, key = "8c56cf43c5629fdaac42e4ae791a317b")
library(newsmap)
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "replValueSp"; definition not updated
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "xMatrix"; definition not updated
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "mMatrix"; definition not updated
library(quanteda)
uk <- mp_availability(countryname == "United Kingdom")</pre>
## Connecting to Manifesto Project DB API...
## Connecting to Manifesto Project DB API... corpus version: 2023-1
## Connecting to Manifesto Project DB API...
## Connecting to Manifesto Project DB API... corpus version: 2023-1
## Connecting to Manifesto Project DB API... corpus version: 2023-1
uk
```

```
##
            Queried for
                               Corpus Version
                                                    Documents found
##
                     98
                                       2023-1
                                                       71 (72.449%)
## Coded Documents found
                              Originals found
                                                         Languages
                                                      1 (english)
           34 (34.694%)
                                 86 (87.755%)
uk<- as.data.frame(uk)</pre>
str(uk)
## 'data.frame':
                   98 obs. of 6 variables:
## $ party : num 51320 51420 51620 51320 51420 ...
## $ date
                : num 194507 194507 194507 195002 195002 ...
## $ manifestos : logi FALSE FALSE FALSE FALSE FALSE ...
## $ originals : logi TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ annotations: logi NA NA NA NA NA NA ...
## $ language
               : chr NA NA NA NA ...
## - attr(*, "query")= MnfstMtd [98 x 15] (S3:
ManifestoMetadata/tbl df/tbl/data.frame)
                                   : num [1:98] 51320 51420 51620 51320
##
     ..$ party
51420 ...
                                   : num [1:98] 194507 194507 194507 195002
##
    ..$ date
195002 ...
     ..$ language
##
                                   : chr [1:98] NA NA NA NA ...
##
     ..$ source
                                   : chr [1:98] NA NA NA NA ...
     ..$ has eu code
                                   : logi [1:98] FALSE FALSE FALSE
##
FALSE FALSE ...
##
     ..$ is_primary_doc
                                   : logi [1:98] NA NA NA NA NA NA ...
     ..$ may contradict core dataset: logi [1:98] NA NA NA NA NA NA ...
##
     ..$ manifesto_id
                                  : chr [1:98] NA NA NA NA ...
##
     ..$ md5sum_text
##
                                   : chr [1:98] NA NA NA NA ...
     ..$ url original
                                   : chr [1:98] "/down/originals/2023-
1/51320_1945.pdf" "/down/originals/2023-1/51420_1945.pdf"
"/down/originals/2023-1/51620_1945.pdf" "/down/originals/2023-
1/51320 1950.pdf" ...
    ..$ md5sum_original
                                  : chr [1:98]
"9ed092419212878b34a602c0014d7c32" "634886d4820ad2b896c15527a5ed7f2d"
"fea8e99005f272abb93f32e1745e7621" "c50581f9acb353a5ca463ff373c4734e" ...
     ..$ annotations
                                  : logi [1:98] NA NA NA NA NA NA ...
##
     ..$ handbook
                                   : chr [1:98] NA NA NA NA ...
     ..$ is copy of
                                   : chr [1:98] NA NA NA NA ...
                                   : chr [1:98] "Let us face the future: a
##
     ..$ title
declaration of Labour policy for the consideration of the nation" "20-point
manifesto of the Liberal Party" "Mr Churchill's declaration of policy to the
electors" "Let us win through together: a declaration of Labour policy for
the consideration of the nation" ...
## - attr(*, "date")= chr "Sat Oct 28 10:08:17 2023"
## - attr(*, "corpus_version")= chr "2023-1"
# Let's focus on the British manifestos of the 2019 elections
uk19 <- mp_corpus(countryname=="United Kingdom" & date == 201912)</pre>
## Connecting to Manifesto Project DB API... corpus version: 2023-1
```

```
uk19
## <<ManifestoCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 10
summary(uk19)
                Length Class
                                         Mode
##
## 51110 201912 1380
                      ManifestoDocument list
## 51210 201912 136
                      ManifestoDocument list
                      ManifestoDocument list
## 51320 201912 1810
## 51340 201912 539
                      ManifestoDocument list
## 51421 201912 1552
                      ManifestoDocument list
## 51430 201912 680
                      ManifestoDocument list
## 51620 201912 1299
                      ManifestoDocument list
                      ManifestoDocument list
## 51901 201912 1081
## 51902 201912 1233
                      ManifestoDocument list
## 51903_201912 617
                      ManifestoDocument list
head(content(uk19[["51110_201912"]]))
## [1] "Foreword"
## [2] "The time to vote Green is now."
## [3] "We know these are dark days."
## [4] "The threat of Brexit hangs over us"
## [5] "and our democracy is under attack."
## [6] "Above all, the climate and environmental emergency rages from the
Amazon to the Arctic."
cmp <- mp maindataset() # Let's download the CMP core dataset</pre>
View(cmp)
uk_cmp <- cmp[ which(cmp$countryname=="United Kingdom" & cmp$date==201912),]
# select name of country and elections
print(uk_cmp [c("partyname", "party", "edate", "date")])
## # A tibble: 10 × 4
##
     partyname
                                         party edate
                                                            date
##
                                         <dbl> <date>
     <chr>>
                                                           <dbl>
## 1 Green Party of England and Wales
                                        51110 2019-12-12 201912
## 2 We Ourselves
                                         51210 2019-12-12 201912
## 3 Labour Party
                                         51320 2019-12-12 201912
## 4 Social Democratic and Labour Party 51340 2019-12-12 201912
## 5 Liberal Democrats
                                         51421 2019-12-12 201912
## 6 Alliance Party of Northern Ireland 51430 2019-12-12 201912
## 7 Conservative Party
                                       51620 2019-12-12 201912
## 8 The Party of Wales
                                       51901 2019-12-12 201912
## 9 Scottish National Party
                                       51902 2019-12-12 201912
## 10 Democratic Unionist Party
                                       51903 2019-12-12 201912
# converting the 10 party manifestoes recovered from the CMP dataset to a
Ouanteda corpus
```

```
quanteda_uk_party <- corpus(uk19)</pre>
summary(head(quanteda uk party ))
## Corpus consisting of 6 documents, showing 6 documents:
##
##
     Text Types Tokens Sentences manifesto id party
                                                       date language source
  text1 4320
##
                37876
                              55 51110 201912 51110 201912
                                                             english MARPOR
## text2
            824
                  3381
                               1 51210 201912 51210 201912
                                                             english MARPOR
## text3 4602 43212
                               1 51320 201912 51320 201912
                                                             english MARPOR
                               3 51340 201912 51340 201912
##
   text4 2352 13956
                                                             english MARPOR
## text5 4654 43092
                               5 51421 201912 51421 201912
                                                             english MARPOR
                               1 51430 201912 51430 201912
##
   text6 2692 18498
                                                             english MARPOR
    has eu code is primary doc may contradict core dataset
##
##
          FALSE
                          TRUE
                                                      FALSE
##
                         md5sum text
                                                               url original
    33ea9f1611ba9f4fc9bc98cc6dd5422d /down/originals/2020-1/51110 2019.pdf
##
    c8c32510e9df570a556c59822bc7422c /down/originals/2020-1/51210 2019.pdf
    Oda2149574aOdccO67abaf223e91b9ab /down/originals/2020-1/51320 2019.pdf
##
    1eaab709947b042b7ed1000bbbcfe09d /down/originals/2020-1/51340 2019.pdf
    436ce3b29af9e0b23ef0f0ea543db9c7 /down/originals/2020-1/51421_2019.pdf
    a32d6fbabf19bf6044552da86c8c1e22 /down/originals/2020-1/51430_2019.pdf
##
##
                     md5sum original annotations handbook is copy of
##
    db0d3e814bdbeed362ea81e53066dc37
                                             TRUE
                                                         5
                                                                 <NA>
##
    1201b63b1ef643cc8e178b72e629a10a
                                                         5
                                             TRUE
                                                                 <NA>
    b92e0bf8d1202c474c16f8f3f6b3925a
                                                         5
##
                                             TRUE
                                                                 <NA>
    99d6e035cf66103f57075215f8233a06
                                             TRUE
                                                         5
                                                                 <NA>
                                                         5
   de1d8528b492ef56f25920534310e4d5
                                             TRUE
                                                                 <NA>
##
    a252c96c570d84baf57f0102ff83ac81
                                             TRUE
                                                         5
                                                                 <NA>
##
                                  title
                                                   id
                      If not now, when? 51110_201912
##
##
                         Time for unity 51210 201912
##
             It's time for real change 51320 201912
##
                Stop Boris, stop Brexit 51340_201912
##
    Stop Brexit build a brighter future 51421_201912
##
                         Demand better. 51430_201912
ndoc(quanteda_uk_party )
## [1] 10
# Let's add the party labels to the doc vars!
uk_cmp$partyname
## [1] "Green Party of England and Wales"
                                              "We Ourselves"
## [3] "Labour Party"
                                              "Social Democratic and Labour
```

```
Party"
## [5] "Liberal Democrats"
                                             "Alliance Party of Northern
Ireland"
## [7] "Conservative Party"
                                             "The Party of Wales"
## [9] "Scottish National Party"
                                             "Democratic Unionist Party"
uk_cmp$party
## [1] 51110 51210 51320 51340 51421 51430 51620 51901 51902 51903
quanteda uk party $party2 <-uk cmp$partyname
quanteda_uk_party $party_label<-uk_cmp$party</pre>
# let's also rename the documents in the corpus according to party labels
docnames(quanteda_uk_party) <- uk_cmp$partyname</pre>
summary(head(quanteda_uk_party))
## Corpus consisting of 6 documents, showing 6 documents:
##
##
                                  Text Types Tokens Sentences manifesto_id
party
      Green Party of England and Wales 4320 37876
                                                           55 51110 201912
51110
##
                          We Ourselves
                                         824
                                               3381
                                                            1 51210_201912
51210
##
                          Labour Party 4602 43212
                                                            1 51320 201912
51320
## Social Democratic and Labour Party 2352 13956
                                                            3 51340 201912
51340
##
                     Liberal Democrats 4654 43092
                                                            5 51421 201912
## Alliance Party of Northern Ireland 2692 18498
                                                            1 51430 201912
51430
      date language source has_eu_code is_primary_doc
may_contradict_core_dataset
## 201912 english MARPOR
                                 FALSE
                                                 TRUE
FALSE
##
                         md5sum text
                                                              url original
## 33ea9f1611ba9f4fc9bc98cc6dd5422d /down/originals/2020-1/51110_2019.pdf
## c8c32510e9df570a556c59822bc7422c /down/originals/2020-1/51210 2019.pdf
```

```
0da2149574a0dcc067abaf223e91b9ab /down/originals/2020-1/51320 2019.pdf
   1eaab709947b042b7ed1000bbbcfe09d /down/originals/2020-1/51340 2019.pdf
## 436ce3b29af9e0b23ef0f0ea543db9c7 /down/originals/2020-1/51421_2019.pdf
    a32d6fbabf19bf6044552da86c8c1e22 /down/originals/2020-1/51430 2019.pdf
                     md5sum_original annotations handbook is_copy_of
##
##
    db0d3e814bdbeed362ea81e53066dc37
                                             TRUE
                                                          5
                                                                  <NA>
                                                          5
    1201b63b1ef643cc8e178b72e629a10a
                                             TRUE
                                                                  <NA>
                                                          5
##
    b92e0bf8d1202c474c16f8f3f6b3925a
                                             TRUE
                                                                  <NA>
                                                          5
    99d6e035cf66103f57075215f8233a06
                                             TRUE
                                                                  <NA>
##
    de1d8528b492ef56f25920534310e4d5
                                             TRUE
                                                          5
                                                                  <NA>
    a252c96c570d84baf57f0102ff83ac81
                                                          5
##
                                             TRUE
                                                                  <NA>
##
                                   title
                                                   id
##
                      If not now, when? 51110 201912
                          Time for unity 51210_201912
##
##
             It's time for real change 51320_201912
                Stop Boris, stop Brexit 51340 201912
##
    Stop Brexit build a brighter future 51421_201912
##
##
                         Demand better. 51430 201912
                                 party2 party label
##
##
      Green Party of England and Wales
                                              51110
##
                          We Ourselves
                                              51210
##
                          Labour Party
                                              51320
    Social Democratic and Labour Party
##
                                              51340
##
                     Liberal Democrats
                                              51421
##
   Alliance Party of Northern Ireland
                                              51430
# Let's tokenize the manifestos
tok_uk <- tokens(quanteda_uk_party , remove_punct = TRUE,</pre>
remove_numbers=TRUE, remove_symbols = TRUE, split_hyphens = TRUE,
remove_separators = TRUE, remove_url = TRUE)
tok_uk <- tokens_remove(tok_uk , stopwords("en"))</pre>
tok uk <- tokens wordstem (tok uk )
# computing the DFM
dfm uk <- dfm(tok uk )</pre>
dfm uk <- dfm remove(dfm uk, stopwords('en'), min nchar = 2)</pre>
head(docvars(dfm_uk ))
##
     manifesto_id party
                          date language source has eu_code is_primary_doc
## 1 51110 201912 51110 201912 english MARPOR
                                                       FALSE
                                                                       TRUE
## 2 51210 201912 51210 201912 english MARPOR
                                                       FALSE
                                                                       TRUE
## 3 51320 201912 51320 201912 english MARPOR
                                                       FALSE
                                                                       TRUE
## 4 51340 201912 51340 201912
                                english MARPOR
                                                                       TRUE
                                                       FALSE
## 5 51421 201912 51421 201912
                                english MARPOR
                                                       FALSE
                                                                       TRUE
## 6 51430 201912 51430 201912 english MARPOR
                                                                       TRUE
                                                       FALSE
##
     may contradict core dataset
                                                        md5sum text
## 1
                            FALSE 33ea9f1611ba9f4fc9bc98cc6dd5422d
## 2
                            FALSE c8c32510e9df570a556c59822bc7422c
## 3
                            FALSE 0da2149574a0dcc067abaf223e91b9ab
## 4
                            FALSE 1eaab709947b042b7ed1000bbbcfe09d
                            FALSE 436ce3b29af9e0b23ef0f0ea543db9c7
## 5
```

```
## 6
                           FALSE a32d6fbabf19bf6044552da86c8c1e22
##
                              url original
                                                             md5sum original
## 1 /down/originals/2020-1/51110_2019.pdf db0d3e814bdbeed362ea81e53066dc37
## 2 /down/originals/2020-1/51210 2019.pdf 1201b63b1ef643cc8e178b72e629a10a
## 3 /down/originals/2020-1/51320_2019.pdf b92e0bf8d1202c474c16f8f3f6b3925a
## 4 /down/originals/2020-1/51340_2019.pdf 99d6e035cf66103f57075215f8233a06
## 5 /down/originals/2020-1/51421 2019.pdf de1d8528b492ef56f25920534310e4d5
## 6 /down/originals/2020-1/51430 2019.pdf a252c96c570d84baf57f0102ff83ac81
     annotations handbook is copy of
                                                        If not now, when?
## 1
            TRUE
                        5
                                <NA>
## 2
                        5
            TRUE
                                <NA>
                                                           Time for unity
            TRUE
                        5
                                <NA>
## 3
                                              It's time for real change
                        5
## 4
            TRUE
                                <NA>
                                                  Stop Boris, stop Brexit
## 5
            TRUE
                        5
                                <NA> Stop Brexit build a brighter future
## 6
            TRUE
                        5
                                <NA>
                                                           Demand better.
##
               id
                                               party2 party_label
## 1 51110 201912
                    Green Party of England and Wales
                                                            51110
## 2 51210 201912
                                        We Ourselves
                                                            51210
## 3 51320_201912
                                         Labour Party
                                                            51320
## 4 51340 201912 Social Democratic and Labour Party
                                                            51340
## 5 51421 201912
                                   Liberal Democrats
                                                            51421
## 6 51430_201912 Alliance Party of Northern Ireland
                                                            51430
```

#### newsmap

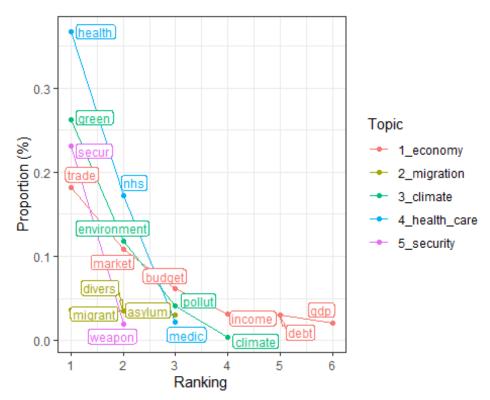
```
# let's create a dictionary of seed-words via the "dictionary" function of
Quanteda and let's apply it to our dfm
dict <- dictionary(list(</pre>
                economy = c("inflation", "econ*", "debt", "trade", "income",
"market", "currency", "gdp", "budget"),
                migration = c("migration", "immigration", "refugee",
"migrant", "asylum", "divers*"),
                climate = c("climate", "warming", "pollut*", "environment",
"green"),
                health care = c("nhs", "health", "medic*"),
                security = c("arms", "secur*", "weapon", "military" )
))
dict
## Dictionary object with 5 key entries.
## - [economy]:
## - inflation, econ*, debt, trade, income, market, currency, gdp, budget
## - [migration]:
     - migration, immigration, refugee, migrant, asylum, divers*
## - [climate]:
     - climate, warming, pollut*, environment, green
## - [health care]:
```

```
## - nhs, health, medic*
## - [security]:
     - arms, secur*, weapon, military
label <- dfm lookup(dfm uk, dictionary = dict)</pre>
label
## Document-feature matrix of: 10 documents, 5 features (4.00% sparse) and 18
docvars.
##
                                         features
## docs
                                          economy migration climate health_care
##
     Green Party of England and Wales
                                              104
                                                         18
                                                                 216
                                                                               57
##
                                                4
                                                          0
                                                                   1
                                                                               2
     We Ourselves
##
     Labour Party
                                              104
                                                         11
                                                                  68
                                                                             111
##
     Social Democratic and Labour Party
                                               46
                                                          1
                                                                  16
                                                                               39
##
     Liberal Democrats
                                                         29
                                                                             169
                                              131
                                                                  46
##
     Alliance Party of Northern Ireland
                                               92
                                                         15
                                                                  16
                                                                              10
##
                                         features
## docs
                                          security
##
     Green Party of England and Wales
                                                41
##
                                                 0
     We Ourselves
##
     Labour Party
                                                69
##
     Social Democratic and Labour Party
                                                16
##
     Liberal Democrats
                                                18
     Alliance Party of Northern Ireland
                                                19
## [ reached max_ndoc ... 4 more documents ]
# let's train the model
model_en <- textmodel_newsmap(dfm_uk, label)</pre>
# now all the words of our texts, including those not included in the seed-
words, get a value!
predict(model_en)
##
     Green Party of England and Wales
                                                              We Ourselves
##
                               climate
                                                                    economy
##
                          Labour Party Social Democratic and Labour Party
##
                           health_care
                                                                health_care
##
                    Liberal Democrats Alliance Party of Northern Ireland
##
                               climate
                                                                    economy
##
                    Conservative Party
                                                        The Party of Wales
##
                               economy
                                                                    climate
##
              Scottish National Party
                                                 Democratic Unionist Party
##
                           health_care
                                                                    climate
## Levels: economy migration climate health care security
```

### **keyATM**

```
library(keyATM)
## keyATM 0.5.0 successfully loaded.
## Papers, examples, resources, and other materials are at
## https://keyatm.github.io/keyATM/
library(ggplot2)
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:NLP':
##
##
       annotate
library(cowplot)
table(ntoken(dfm_uk) > 0)
##
## TRUE
##
     10
keyATM_docs <- keyATM_read(texts = dfm_uk)</pre>
## i Using quanteda dfm.
summary(keyATM_docs)
## keyATM_docs object of 10 documents.
## • Average (min/max) document length: 10654.1 (1215/17418) words
## • Number of unique words: 5786
keywords <- list(</pre>
                economy = c("inflation", "econ", "debt", "trade", "income",
"market", "currency", "gdp", "budget"),
                migration = c("migration", "immigration", "refugee",
"migrant", "asylum", "divers"),
                climate = c("climate", "warming", "pollut", "environment",
"green"),
                health_care = c("nhs", "health", "medic"),
                security = c("arms", "secur", "weapon", "military"))
keywords
```

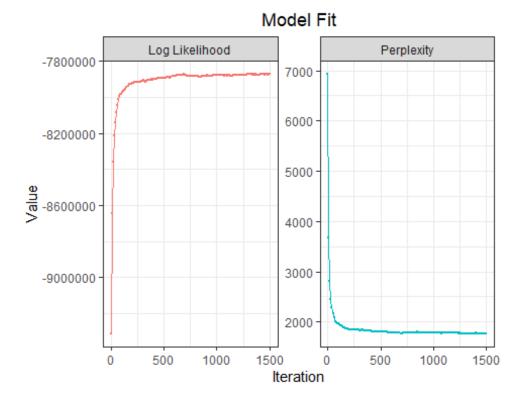
```
## $economy
## [1] "inflation" "econ"
                                "debt"
                                            "trade"
                                                        "income"
                                                                     "market"
## [7] "currency" "gdp"
                                "budget"
##
## $migration
## [1] "migration"
                     "immigration" "refugee"
                                                  "migrant"
                                                                 "asylum"
## [6] "divers"
##
## $climate
## [1] "climate"
                                                  "environment" "green"
                     "warming"
                                    "pollut"
##
## $health care
## [1] "nhs"
                "health" "medic"
##
## $security
## [1] "arms"
                  "secur"
                                         "military"
                              "weapon"
key_viz <- visualize_keywords(docs = keyATM_docs, keywords = keywords)</pre>
## Warning: Keywords are pruned because they do not appear in the documents:
inflation,
## econ, currency, migration, immigration, refugee, warming, arms, and
military
key_viz
```



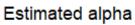
values\_fig(key\_viz)

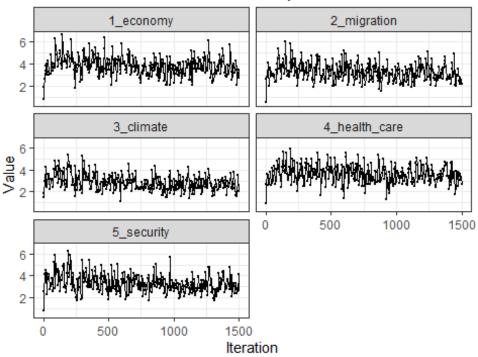
```
## # A tibble: 18 × 5
               Topic [5]
## # Groups:
##
      Word
                  WordCount `Proportion(%)` Ranking Topic
##
                                       <dbl>
                                                <int> <fct>
      <chr>>
                       <int>
## 1 trade
                        194
                                       0.182
                                                    1 1_economy
##
    2 market
                         115
                                       0.108
                                                    2 1_economy
## 3 budget
                                       0.061
                                                    3 1 economy
                         65
## 4 income
                         34
                                       0.032
                                                    4 1_economy
## 5 debt
                         32
                                       0.03
                                                    5 1_economy
## 6 gdp
                         22
                                       0.021
                                                    6 1 economy
## 7 migrant
                         38
                                       0.036
                                                    1 2_migration
                         37
                                                    2 2 migration
## 8 divers
                                       0.035
## 9 asylum
                         32
                                       0.03
                                                    3 2 migration
## 10 green
                         279
                                       0.262
                                                    1 3_climate
## 11 environment
                         126
                                                    2 3_climate
                                       0.118
## 12 pollut
                         44
                                       0.041
                                                    3 3 climate
## 13 climate
                          4
                                       0.004
                                                    4 3 climate
## 14 health
                                                    1 4 health care
                         391
                                       0.367
                                                    2 4 health care
## 15 nhs
                         183
                                       0.172
## 16 medic
                         23
                                       0.022
                                                    3 4 health care
## 17 secur
                         246
                                       0.231
                                                    1 5 security
## 18 weapon
                         21
                                       0.02
                                                    2 5_security
system.time(out <- keyATM(docs = keyATM docs,</pre>
              no_keyword_topics = 1,
              keywords
                                 = keywords,
              model
                                 = "base",
              options
                                 = list(seed = 123)))
## " Initializing the model
## Warning: Keywords are pruned because they do not appear in the documents:
inflation,
## econ, currency, migration, immigration, refugee, warming, arms, and
military
##
                           3bu info@mydup.com
## 1 economy
                 4.301730e-05
                                 4.301730e-05
## 2 migration
                 6.890307e-07
                                 6.890307e-07
## 3 climate
                 6.025015e-07
                                 6.025015e-07
## 4_health_care 4.123490e-07
                                 4.123490e-07
## 5 security
                 6.236813e-07
                                 6.236813e-07
## Other_1
                 8.317489e-07
                                 8.317489e-07
plot_modelfit(out) # If the model is working as expected, we would observe an
```

increase trend for the log-likelihood and a decrease trend for the perplexity



## plot\_alpha(out)

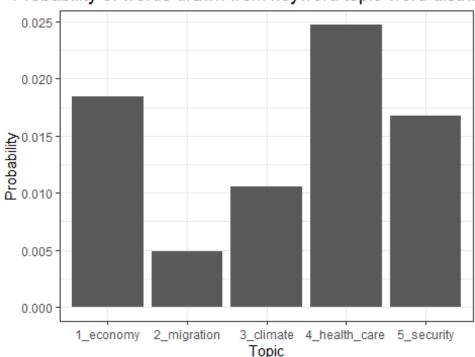




plot\_pi(out) #keywords for migration are not representative thus better be replaced with more frequently appearing keywords. Keywords for climate topic can be reconsidered as well.

## i Plotting pi from the final MCMC draw. Please set `store\_pi` to `TRUE` if you want to plot pi over iterations.

### Probability of words drawn from keyword topic-word distrib



# Adding Covariate(s): Plitical Left-Right dimension applied

```
cmp<-mp maindataset()</pre>
dimension <- cmp$rile</pre>
system.time(out <- keyATM(docs</pre>
                                            = keyATM_docs,
              no keyword topics = 1,
              keywords = keywords,
model = "covariate"
              model
                             = "covariates",
              model settings = list(covariates data = dimension ,
                                         covariates_formula = ~ dimension ),
                                 = list(seed = 123),
              options
                                   = c("Z", "S") ))
                keep
## i Convert covariates data using `model_settings$covariates_formula`.
## " Initializing the model
```

```
## user system elapsed
## 104.57 17.69 125.20
covariates_info(out)
## Colnames: (Intercept), dimension
## Standardization: non-factor
## Formula: ~ dimension
##
## Preview:
## (Intercept) dimension
## 1
           1
                   9.60
            1
                 -37.80
## 2
## 3
                  9.50
           1
## 4
                  28.00
## 5
            1
                 23.81
## 6
             1
                 -44.00
```

#parties are placed along the Political Left-Right dimension: 2 and 6 are being extreme Leftist, 4 and 6 are extreme Rightist, and 1 and 3 are moderate Centre-Rightist