

Text Classification (Newspaper) using Naive Bayes

Data Collection

```
library(jsonlite)
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

library(kableExtra)
api = 'https://content.guardianapis.com/search?'
api_key = '0e9a8c4e-1866-4503-a9eb-70e391c499c4'
page_size = '200'
pages = c(1:5)
queries = c('business','sports','entertainment','economy','politics','science')

news_data = data.frame()
data_count = list()
for(query in queries){
  for (page in pages){
    url = paste(api, 'q=', query, '&page-size=', page_size, '&page=', page, '&api-key=', api_key,
               sep = "")
    json = fromJSON(url)
    data = as.data.frame(json$response$results)
    data_count = append(data_count,nrow(data))
    news_data = rbind(news_data,data)
  }
  print(paste("dfgdf",data_count))

  select_data=select(data,c("sectionName","webPublicationDate","webTitle"))
  print(kable(head(select_data), format = "latex", booktabs = T,
                  caption=paste("Table containing",toupper(query)," News")) %>%
        kable_styling(latex_options = c("striped","hold_position")))
  cat("\n")
  print(paste("Number of rows in ",toupper(query)," News: ",nrow(news_data)))
  print(paste("Number of rows 3in ",toupper(query)," News: ",nrow(data)))
}
```

[1] "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200"

[1] "Number of rows in BUSINESS News: 1000" [1] "Number of rows 3in BUSINESS News: 200" [1] "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" [6] "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200"

[1] "Number of rows in SPORTS News: 2000" [1] "Number of rows 3in SPORTS News: 200" [1] "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" [6] "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200" "dfgdf 200"

Table 1: Table containing BUSINESS News

sectionName	webPublicationDate	webTitle
icare: Workplace safety	2018-06-06T23:58:17Z	How to make friends and love your work
Opinion	2018-04-01T15:38:59Z	Atheists who bring logic to the Easter story are missing the point Juli
Cities	2018-06-06T10:00:10Z	Smoked on the water: the eel trader reviving an old Amsterdam traditio
World news	2018-06-06T06:00:06Z	All eyes on Canada as first G7 nation prepares to make marijuana legal
Education	2017-03-30T06:30:19Z	Business partnership category: award winner and runners up
Business	2018-03-05T21:08:56Z	WTO and top Republican urge Donald Trump not to launch trade war

Table 2: Table containing SPORTS News

sectionName	webPublicationDate	webTitle
Sport	2017-08-17T21:02:33Z	Kevin Durant spurns Trump as teams vow to fund Confederate statue removal
Sport	2017-12-29T16:28:00Z	From Korea to Russia, 2018 set to show again that sport is politics by other mea
Global	2018-10-02T10:59:23Z	Good to meet you ... Ian and Brenda Darke
Opinion	2017-11-27T20:44:30Z	Don't get too excited for Meghan Markle. The British monarchy is oppressive
Sport	2017-12-30T21:00:11Z	Australia's sporting year in review – success stories and slip-ups of 2017 Paul C
Australia news	2018-06-11T18:00:27Z	Taxpayers will need to subsidise Sydney stadium overhauls, documents reveal

200” [11] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”

Table 3: Table containing ENTERTAINMENT News

sectionName	webPublicationDate	webTitle
Football	2018-05-17T13:28:04Z	Mo Salah scores again as boots enter British Museum collection
Life and style	2018-02-05T10:21:24Z	How was your weekend running?
Art and design	2018-08-21T15:30:54Z	‘A menace to life and limb’: the artworks that have injured – and killed
Film	2017-07-09T07:00:18Z	Elle; Certain Women; Neruda; Frantz and more – reviews
Opinion	2018-08-21T12:18:38Z	This beauty pageant is a throwback. It has no place in modern Ireland Tar
Television & radio	2017-01-26T11:56:29Z	Studio Ghibli's first TV show is a wondrous world of peril and magic

[1] “Number of rows in ENTERTAINMENT News: 3000” [1] “Number of rows 3in ENTERTAINMENT News: 200” [1] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [6] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [11] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [16] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”

Table 4: Table containing ECONOMY News

sectionName	webPublicationDate	webTitle
Australia news	2017-10-23T17:00:23Z	Scott Morrison flags health, education and energy reforms
Opinion	2018-08-03T13:15:31Z	Raising interest rates is a big mistake. This will have to be reversed David Bla
Business	2018-04-06T15:46:41Z	Mark Carney warns of climate change threat to financial system
Environment	2018-06-16T06:30:38Z	UK cycling is worth more than the steel industry – where's the strategy?
World news	2018-04-10T16:10:15Z	Tackling the gender pay gap requires more than just lip service Letters
Opinion	2018-09-06T11:34:16Z	Labour's sole focus should be closing Britain's wealth gap Polly Toynbee

[1] “Number of rows in ECONOMY News: 4000” [1] “Number of rows 3in ECONOMY News: 200” [1] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [6] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”

“dfgdf 200” [11] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [16] “dfgdf 200” “dfgdf 200”
“dfgdf 200” “dfgdf 200” “dfgdf 200” [21] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”

Table 5: Table containing POLITICS News

sectionName	webPublicationDate	webTitle
Stage	2018-03-15T12:34:50Z	The best political plays – picked by David Hare, James Graham and more
Politics	2018-04-30T09:38:01Z	Amber Rudd resigns hours after Guardian publishes deportation targets letter
Opinion	2018-08-21T17:42:40Z	The Guardian view on Tory party entryism: a real and present rightwing danger
Politics	2017-11-06T17:43:40Z	Priti Patel reprimanded in person by Theresa May over Israel visit, No 10 reveals
Opinion	2018-03-15T18:50:49Z	The Guardian view on air pollution: moral pusillanimity, political ineptitude Edi
Politics	2018-06-13T20:43:23Z	Shadow minister and five Labour PPSs resign to defy whip over Brexit – as it hap

[1] “Number of rows in POLITICS News: 5000” [1] “Number of rows 3in POLITICS News: 200” [1] “dfgdf
200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [6] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”
“dfgdf 200” [11] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [16] “dfgdf 200” “dfgdf 200”
“dfgdf 200” “dfgdf 200” “dfgdf 200” [21] “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” [26]
“dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200” “dfgdf 200”

Table 6: Table containing SCIENCE News

sectionName	webPublicationDate	webTitle
Opinion	2017-10-02T02:18:49Z	Attacks on the credibility of abuse survivors are not justified by research Micha
Science	2016-12-07T09:31:52Z	Science Museum’s maths gallery soars with stunning Zaha Hadid design
Science	2018-03-26T06:10:22Z	Can you solve it? The art of the deal
Australia news	2018-10-09T04:25:22Z	Wentworth byelection forum: the ‘good-hearted people’ and the missing Liberal
Society	2018-08-03T11:38:54Z	Has your trust in the government and media changed? Share your views
Science	2017-09-24T11:37:58Z	UK invests £65m in Deep Underground Neutrino Experiment in US

[1] “Number of rows in SCIENCE News: 6000” [1] “Number of rows 3in SCIENCE News: 200”

```
summary(news_data)
```

```
##      id              type      sectionId
## Length:6000      Length:6000      Length:6000
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##  sectionName      webPublicationDate  webTitle
## Length:6000      Length:6000      Length:6000
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##   webUrl          apiUrl          isHosted      pillarId
## Length:6000      Length:6000      Mode :logical Length:6000
## Class :character Class :character FALSE:6000      Class :character
## Mode  :character Mode  :character              Mode  :character
##  pillarName
## Length:6000
## Class :character
## Mode  :character
```