Text Classification (Newspaper) using Naive Bayes

Data Collection

```r
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
```

```r
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()
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)
    news_data = rbind(news_data,data)
  }
  print(kable(head(data), format = "latex", booktabs = T,
        caption=paste("Table containing ",toupper(query)," News")) %>%
    kable_styling(latex_options = c("striped","hold_position")))
  cat("\n")
}
```

Table 1: Table containing BUSINESS News

id	type
icare-workplace-safety/2018/jun/07/how-to-make-friends-and-love-your-work	article
commentisfree/2018/apr/01/atheists-logic-easter-faith-belief	article
cities/2018/jun/06/smoked-on-the-water-the-eel-trader-reviving-an-old-amsterdam-tradition	article
world/2018/jun/06/all-eyes-on-canada-as-first-g7-nation-prepares-to-make-marijuana-legal	article
higher-education-network/2017/mar/30/business-partnership-category-award-winner-and-runners-up	article
business/live/2018/mar/05/trade-war-fears-italian-election-results-markets-euro-services-car-sales-business-live	liveblog

```r
summary(news_data)
```

```
##       id                type              sectionId
```

Table 2: Table containing SPORTS News

id	type	sectio...
sport/2017/aug/17/kevin-durant-white-house-visit-donald-trump	article	sport
sport/blog/2017/dec/29/korea-russia-sport-politics-andy-bull	article	sport
global/2018/oct/02/good-to-meet-you-ian-and-brenda-darke	article	global
sport/2017/dec/31/australian-sporting-year-in-review-2017	article	sport
commentisfree/2017/nov/27/meghan-markle-british-monarchy-oppressive	article	comm
australia-news/2018/jun/12/taxpayers-will-need-to-subsidise-sydney-stadium-overhauls-documents-reveal	article	austra

Table 3: Table containing ENTERTAINMENT News

id	type	section
tv-and-radio/2018/aug/20/janet-hargreaves-obituary	article	tv-and
music/2018/feb/18/morrissey-review-hydro-glasgow	article	music
lifeandstyle/2016/dec/24/how-victorian-festivities-began-to-go-with-a-bang-christmas-cracker-tom-smith	article	lifeand
artanddesign/2018/apr/09/diane-arbus-smithsonian-museum-box-of-ten-photographs	article	artand
universal-sony-pictures-home-entertainment-victoria-and-abdul/2017/dec/11/the-real-victoria-and-abdul	article	univers
global/2018/sep/23/get-your-skates-on-the-rise-of-british-ice-hockey	article	sport

Table 4: Table containing ECONOMY News

id	type	sectionId	sectionN
australia-news/2017/oct/24/scott-morrison-flags-health-education-and-energy-reforms	article	australia-news	Australi
commentisfree/2018/aug/03/raising-interest-rates-mpc-wage-growth	article	commentisfree	Opinion
business/2018/apr/06/mark-carney-warns-climate-change-threat-financial-system	article	business	Business
environment/bike-blog/2018/jun/16/uk-cycling-steel-industry-strategy	article	environment	Environ
news/2018/apr/10/tackling-the-gender-pay-gap-requires-more-than-just-lip-service	article	world	World n
commentisfree/2018/sep/06/labour-focus-closing-britain-wealth-gap-super-rich	article	commentisfree	Opinion

Table 5: Table containing POLITICS News

id
politics/2018/apr/29/amber-rudd-letter-to-pm-reveals-ambitious-but-deliverable-removals-target
politics/2018/apr/29/amber-rudd-resigns-as-home-secretary-after-windrush-scandal
politics/blog/live/2017/apr/11/politics-live-readers-edition-tuesday-11-april
politics/blog/live/2017/apr/13/politics-live-readers-edition-thursday-13-april
politics/blog/live/2018/jun/13/brexit-parliament-wont-be-allowed-to-take-charge-of-talks-in-no-deal-scenario-minister-tells-
commentisfree/2018/mar/15/the-guardian-view-on-air-pollution-moral-pusillanimity-political-ineptitude

```
##  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
```

Table 6: Table containing SCIENCE News

id
commentisfree/2017/oct/02/attacks-on-the-credibility-of-abuse-survivors-are-not-justified-by-research
science/alexs-adventures-in-numberland/2016/dec/07/science-museums-maths-gallery-soars-with-stunning-zaha-hadid-desi
science/2018/mar/26/can-you-solve-it-the-art-of-the-deal
australia-news/2018/oct/09/wentworth-byelection-forum-the-good-hearted-people-and-the-missing-liberal
society/2018/aug/03/trust-government-media-confident-banks-falling
science/life-and-physics/2017/sep/24/uk-invests-65m-in-deep-underground-neutrino-experiment-in-us

```
##  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
```