

RiBD Praca Domowa 1

Grzegorz Bajor

6 marca 2016

```
library(stringi)
```

Na podstawie historii tweetów z dni 01.03.2016, 29.02.2016, 28.02.2016 o Oscarach, (to jest z dnia przed i dwóch kolejnych dni po gali) (po 50 000 tweetów z każdego dnia) przeanalizujemy o jakich filmach i osobach najczęściej mówiono w tweetach.

```
colnames(tw)
```

```
## [1] "X"           "text"         "favorited"    "favoriteCount"
## [5] "replyToSN"   "created"      "truncated"    "replyToSID"
## [9] "id"          "replyToUID"   "statusSource" "screenName"
## [13] "retweetCount" "isRetweet"    "retweeted"    "longitude"
## [17] "latitude"
```

```
#wczytywanie treści tweetów z kolejnych dni
```

```
twee <- vector("list", 3)
```

```
twee[[1]] <- as.list(levels(tw[[2]]))
twee[[2]] <- as.list(levels(tw2[[2]]))
twee[[3]] <- as.list(levels(tw3[[2]]))
```

```
#zliczanie w ilu różnych tweetach pojawiają się szukane hasła
```

```
for(i in 1:3){
  leo[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)Leonardo|di( ?)caprio") != 0)
  rylance[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)mark|rylance") != 0)
  brie[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)brie|larson") != 0)
  alicia[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)alicia|vikander") != 0)
  sam[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)sam|smith") != 0)
  ennio[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)ennio|morricone") != 0)

  amy[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)amy") != 0)
  river[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)river") != 0)
  stut[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)stutterer") != 0)
  bear[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)bear") != 0)
  madmax[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)madmax|mad max") != 0)
  machina[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)machina") != 0)
  spectre[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)spectre") != 0)
  spotlight[i] <- sum(sapply(twee[[i]], stri_count_regex, "(?i)spotlight") != 0)
}
```

```
leo #Leonardo DiCaprio
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      4659      5181      4458
```

Jeśli chodzi o DiCaprio, to liczby tweetów nie wydają się być zaskakujące dla którego to była już piąta nominacja do Oscara.

Jeszcze dla potwierdzenia wyekstrahujemy wszystkie wyrazy z dnia 01.03.2016

```
toLow <- stri_trans_tolower(twee[[1]])
x <- table(unlist(stri_extract_all_regex(toLow, "\\w+")))
which(x > 1000)
```

```
##      00a0      00b8      00bd      2016      a      an      and      at
##      155      180      185      1113     4325     6223     6290     7430
##      co      de      del dicaprio      do      e      ed      el
##      12787    15132    15427    16119    16720    17491    17746    18078
##      en      for      https      i      in      is      la      leo
##      18531    21441    25677    25956    26657    27469    30641    31273
## leonardo    los      me      no      o      of      oscar    oscars
##      31307    32203    33813    36796    37427    37743    38559    38749
##      óscar    para    party    por      que      rt      s      se
##      39108    39445    39603    41233    43044    45571    45801    46637
##      su      t      the      to      u      um      un      won
##      49119    49638    50628    51295    52379    52770    52806    55757
##      y      you    youtube
##      56733    57318    57340
```

I można zobaczyć, że oprócz przymków i spójników najczęściej pojawiają się hasła dotyczące Leonarda DiCaprio.

Pozostałe osoby, które otrzymały nagrody:

```
brie #Brie Larson
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      298      252      217
```

```
rylance #Mark Rylance
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      169      192      89
```

```
alicia #Alicia Vikander
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      110      92      230
```

```
sam #Sam Smith
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      584      631      299
```

```
ennio #Ennio Morricone
```

```
## 03.01.2016 02.29.2016 02.28.2016
##      292      220      40
```

i filmy:

amy *#Amy*

##	03.01.2016	02.29.2016	02.28.2016
##	200	174	39

river *#A girl in the river*

##	03.01.2016	02.29.2016	02.28.2016
##	80	73	18

stut *#Stutterer*

##	03.01.2016	02.29.2016	02.28.2016
##	20	90	2

bear *#Bear story*

##	03.01.2016	02.29.2016	02.28.2016
##	219	167	87

madmax *#Mad max*

##	03.01.2016	02.29.2016	02.28.2016
##	165	202	165

machina *#Ex machina*

##	03.01.2016	02.29.2016	02.28.2016
##	76	45	20

spectre *#Spectre*

##	03.01.2016	02.29.2016	02.28.2016
##	6	15	9

spotlight *#Spotlight*

##	03.01.2016	02.29.2016	02.28.2016
##	521	548	134

Dla pozostałych osób wyniki oscylują w granicy od 100 do 600 tweetów.

Na podstawie wyników można stwierdzić, że przed rozdaniem nagród, ludzie nieprzewidywali niektórych osób (Mark Rylance, Ennio Morricone) i filmów (Ex Machina, Stutterer, Amy) do otrzymania statuetki.