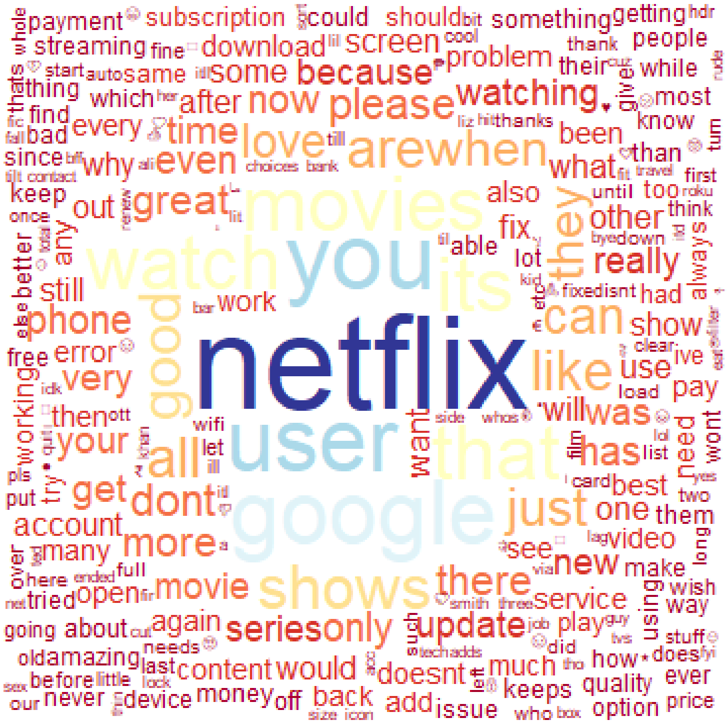
# NLP_Netflix_Reviews

Nabil Momin

2024-06-17

```
library(corrgram)
library(corrplot)

## corrplot 0.92 loaded

library(caTools)
library(Amelia)

## Loading required package: Rcpp

## ## 
## ## Amelia II: Multiple Imputation
## ## (Version 1.8.2, built: 2024-04-10)
## ## Copyright (C) 2005-2024 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##

library(ggplot2)
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(rpart)
library(rpart.plot)
library(randomForest)

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

## 
## Attaching package: 'randomForest'
## The following object is masked from 'package:dplyr':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin

library(ISLR)
library(e1071)
library(cluster)
library(tm)

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##     annotate

library(twitteR)

##
## Attaching package: 'twitteR'

## The following objects are masked from 'package:dplyr':
##
##     id, location

library(wordcloud)

## Loading required package: RColorBrewer

library(RColorBrewer)

#### getting the data into text

netflix <- read.csv('netflix_reviews.csv')

#### text mining

docs <- Corpus(VectorSource(netflix))

#### using TermDocumentMatrix to do manipulation

doc.matrix <- TermDocumentMatrix(docs,control=list(removePunctuation = TRUE,

stopwords=c('the','The','and','build','but','app','not','have','cant','with',
'for','this','from'),
                                                   removeNumbers = TRUE,
                                                       tolower = TRUE))


#### make it a matrix

doc.matrix <- as.matrix(doc.matrix)

#### word count

word.freq <- sort(rowSums(doc.matrix),decreasing = TRUE)

df <- data.frame(word=names(word.freq),freq = word.freq)

#### wordcloud NLP

wordcloud(df$word,df$freq,random.order = F, colors = brewer.pal(11,
'RdYlBu'))
```


whole
payment
subscription
could
should
bit
something
getting
hdr
streaming
fine
download
screen
cool
problem
thank
people
start
auto
same
some
because
their
while
rude
thats
thing
which
her
after
now
please
watching
give
most
find
bad
every
time
love
till
are
when
thanks
been
know
turn
since
why
even
choices
bank
what
than
first
contact
keep
out
great
movies
also
until
too
roku
once
other
think
better
any
watch
you
its
able
fix
they
down
really
always
still
work
lot
kid
fixed
isnt
had
else
phone
can
show
good
clear
free
error
netflix
like
ive
very
idk
use
load
pay
working
then
ott
wifi
will
was
your
let
want
side
whos
film
need
wont
all
user
that
has
list
pls
put
try
get
dont
google
card
best
yes
just
one
two
them
account
more
video
long
over
many
see
new
make
here
ended
full
via
tried
open
movie
shows
there
wish
way
using
service
smith
three
going
about
cut
again
series
only
update
job
play
guy
needs
such
did
stuff
old
amazing
tech
adds
does
last
content
would
doesnt
much
how
before
little
lock
ever
keeps
quality
our
never
device
money
off
back
add
issue
left
who
box
option
price
size
icon