test

2022-11-06

```
#test4
# df = read.csv("amazon_review_polarity_csv/train.csv", header = FALSE)
N = 100000
N t = 0.8*N
reviews_text<-readLines("amazon_review_polarity_csv/train.csv", n = N)
reviews_text<-data.frame(reviews_text)</pre>
library(tidyr)
reviews_text<-separate(data = reviews_text, col = reviews_text, into = c("Sentiment", "SentimentText"),
# Retaining only alphanumeric values in the sentiment column
reviews_text$Sentiment<-gsub("[^[:alnum:]]","",reviews_text$Sentiment)
# Retaining only alphanumeric values in the sentiment text
reviews_text$SentimentText<-gsub("[^[:alnum:]]"," ",reviews_text$SentimentText)</pre>
# Replacing multiple spaces in the text with single space
reviews\_text\$SentimentText <-gsub("(?<=[\\s])\\s*|^\s+|\\s+|", "", reviews\_text\$SentimentText, perl=TRU, reviews\_text\$SentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSentimentSe
# Writing the output to a file that can be consumed in other projects
write.table(reviews_text,file = "Sentiment Analysis Dataset.csv",row.names = F,col.names = T,sep=',')
# reading the first 1000 reviews from the dataset
reviews_text<-readLines('amazon_review_polarity_csv/train.csv', n = N)</pre>
# basic EDA to confirm that the data is read correctly
print(class(reviews_text))
## [1] "character"
print(length(reviews_text))
## [1] 100000
print(head(reviews_text,2))
\# [1] "\"2\",\"Stuning even for the non-gamer\",\"This sound track was beautiful! It paints the senery
## [2] "\"2\",\"The best soundtrack ever to anything.\",\"I'm reading a lot of reviews saying that this
# replacing the positive sentiment value 2 with __label__2
reviews_text<-gsub("\\\",",","__label__2 ",reviews_text)</pre>
# replacing the negative sentiment value 1 with __label__1
```

reviews_text<-gsub("\\\",",","__label__1 ",reviews_text)</pre>

```
# removing the unnecessary \" characters
reviews_text<-gsub("\\\""," ",reviews_text)</pre>
# replacing multiple spaces in the text with single space
reviews\_text <-gsub("(?<=[\s])\s*|^\s+|\s+$", "", reviews\_text, perl=TRUE)
# Basic EDA post the required processing to confirm input is as desired
print("EDA POST PROCESSING")
## [1] "EDA POST PROCESSING"
print(class(reviews_text))
## [1] "character"
print(length(reviews_text))
## [1] 100000
print(head(reviews_text,2))
## [1] "__label__2 Stuning even for the non-gamer , This sound track was beautiful! It paints the sener
## [2] "__label__2 The best soundtrack ever to anything. , I'm reading a lot of reviews saying that thi
# writing the revamped file to the directory so we could use it with
# fastText sentiment analyzer project
fileConn<-file("Sentiment Analysis Dataset_ft.txt")</pre>
writeLines(reviews text, fileConn)
close(fileConn)
BoW approach
# including the required libraries
library(SnowballC)
library(tm)
## Loading required package: NLP
# setting the working directory where the text reviews dataset is located
# reading the transformed file as a dataframe
text <- read.table(file='Sentiment Analysis Dataset.csv', sep=',', header = TRUE)
# checking the dataframe to confirm everything is in tact
print(dim(text))
## [1] 100000
                   2
# View(text)
```

```
# transforming the text into volatile corpus
train_corp = VCorpus(VectorSource(text$SentimentText))
print(train corp)
## <<VCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 100000
# creating document term matrix
dtm_train <- DocumentTermMatrix(train_corp, control = list( tolower = TRUE, removeNumbers = TRUE, stopwo
# Basic EDA on dtm
inspect(dtm_train)
## <<DocumentTermMatrix (documents: 100000, terms: 74760)>>
## Non-/sparse entries: 3399444/7472600556
## Sparsity
                  : 100%
## Maximal term length: 188
## Weighting
              : term frequency (tf)
## Sample
##
        Terms
## Docs
         book get good great just like movi one read time
##
    1250
            0
                   0
                         0
                              0
                                  0
                                          0
               0
                                       0
##
    56817
            0 0
                    0
                         0
                                                    0
##
    63995
            0
               2
                              0
                                  2
                                          2
                    1
                         1
                                       0
                                               0
                                                    1
##
    6785
            0
               7
                    1
                         0
                              0
                                                    0
##
    69262
          0 0 0
                         0
                            0 0
                                     0 0
                                                   0
##
    73633
          1 0 0
                         2 0 3
                                                   2
##
    79144
          0 0 0
                              0 0
                                     0 0 0
                         0
                                                   0
          0 0
##
    80872
                  0
                         0
                              0
                                                   0
            0 1 0
                         0
##
    85894
                              0 1
                                       0 1 0
                                                   0
##
    87875
# Removing sparse terms
dtm_train = removeSparseTerms(dtm_train, 0.99)
inspect(dtm_train)
## <<DocumentTermMatrix (documents: 100000, terms: 645)>>
## Non-/sparse entries: 2131029/62368971
## Sparsity
                  : 97%
## Maximal term length: 10
## Weighting
                 : term frequency (tf)
## Sample
                   :
##
## Docs
         book get good great just like movi one read time
##
    34297
            0
               1
                    0
                         0
                              2
                                  5
                                           3
##
    38984
            6 0
                  0
                              1
                                  0
                                          1
                                               0
                         1
                                       0
                                                    1
##
    42051
           3 0 1
                         0
                            2
                                         3
##
    56269
          0 1 0
                         0
                            1
                                  0
                                       1
                                         1
                                                   1
##
    65117
           0 1 1
                         1
                              1
                                  0
                                       0 1
                                               0
                                                   0
##
          0 0 1
                         2 0 4
                                     0 0 1
                                                   0
    65135
##
    6785
          0 7 1
                         0 0 1
```

0

##

80366

0 3 1

1

1

1

```
##
   87149
         6 0 1
                    0
                      0 1
##
   90397
# splitting the train and test DTM
dtm_train_train <- dtm_train[1:N_t, ]</pre>
dtm train test <- dtm train[(N t+1):N, ]</pre>
dtm_train_train_labels <- as.factor(as.character(text[1:N_t, ]$Sentiment))</pre>
dtm_train_test_labels <- as.factor(as.character(text[(N_t+1):N, ]$Sentiment))</pre>
cellconvert<- function(x) { x <- ifelse(x > 0, "Y", "N") }
# applying the function to rows in training and test datasets
dtm_train_train <- apply(dtm_train_train, MARGIN = 2,cellconvert)</pre>
dtm_train_test <- apply(dtm_train_test, MARGIN = 2,cellconvert)</pre>
# inspecting the train dtm to confirm all is in tact
# View(dtm_train_train)
training model
# training the naive bayes classifier on the training dtm
library(e1071)
nb_senti_classifier=naiveBayes(dtm_train_train,dtm_train_train_labels)
# printing the summary of the model created
summary(nb senti classifier)
         Length Class Mode
          2 table numeric
## apriori
## tables
         645
             -none- list
## levels
         2
             -none- character
## isnumeric 645 -none-logical
              -none- call
## call
          3
# making predictions on the test data dtm
nb_predicts <- predict(nb_senti_classifier, dtm_train_test,type="class")</pre>
# printing the predictions from the model
print(nb_predicts)
##
     ##
##
    [73] \ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2
   ##
##
   ##
   ##
##
   [253] 1 2 1 1 2 2 2 2 1 2 1 1 2 2 1 1 2 2 1 2 1 2 1 1 1 1 2 2 2 2 2 2 2 1
##
   ##
   [325] 2 2 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 2 1 2 1 1 1 1 1 2 2 1 2
   ##
##
   [433] 1 2 2 1 2 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 2 2 2 2 1 1 1 1 1 2 1 2 2 2
##
```

[469] 2 2 2 2 2 1 1 2 1 2 1 1 2 2 1 1 2 2 2 1 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 1 ## [505] 2 2 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 2 2 2 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 1 1 1 ## ## ## $[613] \ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2$ ## ## ## ## ## ## ## $[865] \ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2$ ## [901] 2 2 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 1 2 2 2 2 2 1 1 1 1 ## ## ## $\begin{smallmatrix} 1045 \end{smallmatrix}] \ 2\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$ ## ## ## ## [1153] 1 2 2 1 1 1 2 2 2 2 2 2 1 1 2 2 1 1 1 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 ## $[1189] \ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1$ ## ## ## ## ## [1369] 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 1 2 2 1 1 2 1 1 1 2 1 1 2 1 1 2 ## ## ## ## ## $[1549] \ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1$ ## $[1585] \ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2$ ## ## ## ## ## ## [1801] 2 2 2 1 1 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 1 2 2 2 2 2 1 $[1837] \ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2$ ## $[1873] \ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 1$ ## ## ## ## $\begin{smallmatrix} 2089 \end{smallmatrix}] \ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 2\ 1\ 2$ ## ## ## ## $\begin{smallmatrix} 2233 \end{smallmatrix}] \ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2$ ## $[2269] \ 2\ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2$ ## ## ## [2377] 2 1 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1

[6301] 2 1 2 1 2 2 2 2 1 1 2 2 2 2 1 1 1 2 1 1 2 1 1 1 2 1 1 2 2 1 1 1 2 2 1 2 2 2 1 $[6337] \ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1$ ## ## ## ## $[6481] \ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2$ ## [6517] 1 2 2 2 2 2 1 1 2 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1 2 2 2 1 1 2 1 1 1 2 1 1 ## ## $[6589] \ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2$ ## ## ## [6769] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 1 ## ## ## $[6877] \ 2\ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2$ ## ## $[6949] \ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2$ ## ## ## ## ## ## ## ## $[7237] \ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1$ ## ## ## ## [7381] 2 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 1 1 1 2 2 2 1 1 ## ## ## $[7453] \ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2$ $[7489] \ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 1$ ## ## [7525] 2 2 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1 1 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 ## [7561] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 1 1 2 2 2 2 2 2 1 1 1 ## ## ## ## ## ## ## $[7849] \ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2$ ## ## ## ## ## ## ## ## ## ## ## ## ## ## ## ## $[8425] \ 2\ 1\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2\ 1$ ## ## ## [8533] 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 1 1 2 1 2 2 1 1 1 2 2 1 ## ## ## ## ## ## ## $[8821] \ 2\ 1\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1$ ## ## ## ## ## ## ## [9073] 2 2 1 1 2 2 2 2 2 2 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 2 2 1 2 2 2 1 1 2 2 2 ## ## ## ## ## ## ## [9289] 2 1 1 1 2 2 2 1 2 1 2 2 2 2 2 1 2 1 1 1 2 2 1 2 2 2 2 2 1 1 1 2 2 1 1 1 1 ## ## ## ## ## ## ## ## ## ## ## ## ## ## $[9865] \ 2\ 1\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 1\ 1$ ## ## ## ## [10153] 2 2 1 2 1 1 1 2 1 2 2 2 2 2 2 2 1 1 2 1 1 1 1 2 2 2 2 2 1 2 2 2 1 2 2 2 1

```
## [10225] 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 1 2 2 1 1 2 2 1 2 2 2 2 1
## [10261] 2 2 2 2 2 1 2 1 1 1 1 2 2 2 1 1 1 1 2 1 1 1 1 2 2 2 2 1 2 1 2 1 2 1 2
## [10333] 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 2
## [10369] 1 2 2 1 1 1 1 1 1 2 1 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 2 2 1 1 2 1 2 1 1 2 1 1
## [10405] 1 2 1 2 2 2 2 1 1 1 1 1 2 2 1 1 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1
## [10513] 1 2 2 1 2 2 1 1 1 1 1 2 1 2 2 1 1 1 2 2 2 1 2 1 1 1 1 1 1 1 2 1 2 2 2 1
## [10729] 2 1 1 1 2 2 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2
## [10765] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 2 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 2
## [10873] 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 1 1 1 1 1 2 1 2 2 2 2 2 1 1 1 1 1 1 2 1 1
## [10945] 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 2 1 2 2 1 2 2 1 2 1 1 1 2 1 2 1 1 2 1 1
## [10981] 1 1 1 1 2 2 1 2 1 1 2 1 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2
## [11053] 2 1 2 1 1 2 1 2 1 2 1 1 1 1 1 1 1 2 2 2 2 2 1 1 2 1 2 1 1 2 1 1 1 1 1 2 1 2
## [11125] 2 2 2 1 1 2 2 1 2 1 1 2 1 1 1 1 1 1 2 1 2 2 2 2 2 2 2 1 1 1 1 1 2 2 1 1 1 1 2 2
## [11197] 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2 1 2 2 1 1 2
## [11269] 1 1 2 2 2 2 2 2 2 2 2 1 1 1 1 2 1 1 2 1 1 1 2 2 2 2 2 2 1 2 2 1 2 1 1 1 2
## [11305] 1 2 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 2 1 2 2 1 1 1 1 2 2 2 2 2 1 2 2
## [11377] 2 1 1 1 1 1 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 
## [11485] 2 1 1 1 2 2 1 1 2 2 1 2 1 2 2 2 1 1 2 2 2 1 1 1 1 1 1 1 1 2 2 2 2 2
## [11629] 1 2 2 1 1 1 1 1 2 2 1 1 2 2 1 1 1 2 1 1 2 2 1 2 2 1 2 2 2 1 1 2 2 2 1 1 2 1 1
## [11773] 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 1 2 2 2 2 2 2 2 1 2 1 1 2 1 1 1 2 2 2 2 2 1 2
## [11809] 2 1 2 1 1 2 1 2 2 1 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 2 2 2 2 1 1 1 1 1 1 2 2 1
## [11881] 2 1 1 1 2 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2
## [11917] 2 2 2 2 2 1 2 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 2 2 1 1 2 1 1
## [11989] 2 2 2 1 1 2 1 2 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 2 2 1 1 1 2 2 2 2 1
```

```
## [12169] 1 2 2 2 2 1 2 1 1 1 1 2 2 2 2 2 1 1 1 1 1 1 2 2 2 1 2 1 2 1 2 2 2 1 2 1 2 1 2
## [12349] 1 2 2 1 1 1 1 1 2 1 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 2 1 2 1 1 1 2 1 2 2 2
## [12385] 2 2 1 2 2 2 1 2 1 2 1 2 2 2 1 2 1 1 1 1 1 1 1 2 2 1 1 2 1 2 1 2 1 1 1 2 1 1
## [12421] 1 2 1 1 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 2 1
## [12457] 2 1 1 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 2 1 2 1 2 2 2 2 1 1 1 2 1 2 2 1 2 1 2 1 1 1
## [12493] 2 1 1 2 2 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 2 1
## [12529] 2 2 2 1 2 1 1 1 1 2 2 1 1 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 1 2 1 1 1 1 2 1 2 1
## [12637] 2 1 2 2 2 2 2 2 2 2 1 1 1 2 1 1 1 1 2 1 2 2 1 1 1 2 2 1 2 2 2 2 1 1 1 2
## [12709] 1 2 1 1 1 2 2 1 2 1 2 1 2 1 2 1 2 2 2 1 1 2 2 1 2 1 2 1 2 1 2 2 2
## [12745] 2 2 2 1 1 2 2 2 1 2 1 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 1 1 2 1 2
## [12817] 1 2 2 1 2 1 1 1 2 2 2 2 1 1 1 1 2 2 1 1 1 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2 1
## [12853] 1 1 1 1 1 2 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 2 1 2 1 2 1 1 1
## [12889] 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2
## [12925] 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 2 1 1 1 2 1 2 1 1 1 2 1 2 1 2 1 1 1 2 1 2 2 2 2 2
## [13033] 2 2 2 2 2 1 1 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1 2
## [13069] 2 2 2 2 2 2 2 1 1 1 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 2 2 2 1 2 1 2 2 1 2 1 2 2 1 1
## [13285] 2 2 2 2 2 2 1 2 2 2 1 1 2 2 1 1 2 2 2 1 2 1 2 1 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1
## [13393] 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 2 2 2 1 1 1 1 1 2 2 2 2 2 1 2 2 2
## [13501] 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [13537] 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 2 2 2 2 2 2 1 1 2 2 2 1 1 2 2 1 1
## [13573] 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 2 2 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1
## [13717] 2 2 2 2 1 2 2 1 2 1 2 1 2 1 1 2 1 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 1 2 2 1
## [13825] 2 1 1 1 2 2 1 1 1 1 1 2 2 1 2 1 1 2 1 2 2 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 2
## [13933] 2 1 1 1 1 1 1 2 2 2 2 1 1 1 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 1 2 1 1 1 1 2 2 2
```

```
## [14113] 2 2 2 1 1 2 1 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 1 2 1 2
## [14185] 2 2 1 1 2 2 1 2 1 1 1 2 2 1 2 1 1 1 2 2 1 2 1 1 1 2 2 2 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 1 2
## [14293] 1 1 1 2 1 2 1 2 1 1 1 1 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1
## [14365] 2 2 1 2 1 2 2 2 1 1 1 2 2 2 2 2 1 2 2 2 1 2 2 1 1 1 2 2 2 1 1 1 2 2 2 1 1 1
## [14473] 1 1 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 1
## [14725] 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 1 1 1 1 2 1 2 1 2 1 2 2 2 2 2 2
## [14797] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 2 1
## [14869] 2 1 2 2 2 2 2 1 1 2 2 1 1 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1
## [14905] 2 1 1 1 1 2 2 1 2 2 1 2 1 1 1 1 2 1 1 2 2 1 2 2 1 1 1 1 1 2 2 2
## [14941] 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 2 1 1 1 2 2 2 2 1 1 1 2 2 2 1
## [14977] 2 1 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1 2 1 1 1 2 2 2 2 2 2 1 1 1 1 1
## [15013] 2 2 1 1 1 2 2 1 2 1 1 1 2 2 1 2 1 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 2 2 1 2 1 2 1 1 1 2
## [15049] 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 2 2 1 1 2 1 2 1 1 1 1 1
## [15085] 2 2 2 1 2 1 1 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 2 1 2 1 2 2 2 2 1 1 1 2 2 2 2 2 1 1
## [15121] 1 2 2 1 1 2 2 1 2 1 2 2 2 2 2 1 2 2 2 2 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1
## [15157] 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 2 2 2 2 1 1 2 1 2 1 2 1 2 1
## [15193] 1 2 2 2 2 1 2 1 2 1 2 1 1 1 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2
## [15337] 2 2 2 2 1 2 1 1 1 1 1 1 2 1 1 2 2 1 1 1 2 1 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2
## [15373] 2 2 2 2 1 2 2 2 2 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 2 2 2 1 1 2 2 2
## [15409] 2 2 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2
## [15553] 1 2 1 2 1 2 1 1 2 1 2 1 1 1 2 2 2 1 2 2 2 1 2 1 1 1 2 1 2 1 2 1 2 1 2 2 2 1
## [15589] 2 2 2 1 2 2 1 1 1 1 2 1 1 2 1 1 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1
## [15625] 2 2 1 1 2 2 2 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 2 2 1 2 1 2 1 2 1 2 2 2 2 2 1 2
## [15661] 1 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 2 2 2
## [15697] 2 1 2 1 2 2 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 1 2 2 2 2 2 1 1 1 1 2 2 2 2 2 1 1 2 1 2 1 1
## [15805] 1 1 1 2 2 1 2 2 1 2 2 2 2 1 1 2 1 1 2 1 1 2 2 2 2 1 1 1 1 1 2 2 1 2 2 1 2 2
## [15949] 2 2 2 1 1 1 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 1 1 1 2
## [15985] 1 1 2 1 2 1 1 2 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 2 2 1 1 2 2 2 2 2 2
```

```
## [16057] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
## [16237] 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2
## [16273] 1 2 2 2 2 2 1 2 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1
## [16381] 2 2 2 2 2 2 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2 1 1 1 1 1 2 2 2 2 2 1 1 1 1 1 2 1 2 2 1 2 1 1 1
## [16525] 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1 2 1 1 1 2 2 2 2 2 1 1 1 1 2 2
## [16561] 2 1 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 2 2 2 1 1 2 1 2 2 2 1 1 1 1 1 1 1 1
## [16597] 1 1 1 1 1 1 1 2 2 2 2 2 1 1 1 1 2 2 2 1 1 1 1 2 2 2 1 2 1 2 1 2 1 2 1 2 1 1 1 1
## [16669] 2 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1 1 1 2 2 1 1 1 2 2 1 1 1 2 1 1 2 1 1 2 1 2 2 1 2
## [16705] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 2 2 1 1 1 1 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1
## [16849] 2 1 2 2 1 1 2 1 1 2 1 2 2 1 1 1 2 2 1 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2
## [16885] 2 2 1 1 1 2 2 1 1 2 1 1 2 2 2 1 1 2 1 2 2 2 1 1 2 1 2 2 1 2 1 1 1 1 2 2 1 1 2 1 2 1 1
## [16993] 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 1 2 2 2 1 1
## [17029] 1 1 1 2 1 1 2 2 1 2 2 1 2 1 1 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 1 2 1
## [17101] 2 2 1 1 2 2 2 1 1 1 1 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1
## [17173] 1 2 2 2 2 1 1 1 2 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 2 1 2 2
## [17209] 2 2 2 2 1 2 2 1 1 2 2 1 1 1 2 2 1 1 1 1 2 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 2 2 1 2 1 2 1 1 1
## [17281] 2 2 2 2 1 2 2 1 1 2 1 2 2 2 1 1 2 2 2 2 1 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 2 1 1 1
## [17317] 1 2 1 1 2 1 1 2 2 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 1 1 1 1 1 1 1 1 2 2 2 1 2
## [17353] 2 1 1 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1 1 1 2 1 2 1 2 1 2 2 2 2 2 2 2 1 2 1 1
## [17389] 1 1 2 1 1 2 2 2 1 2 2 2 1 1 2 1 1 2 2 2 1 1 1 1 1 2 2 1 2 2 2 1 2 1 2 1 2
## [17425] 1 2 2 2 1 2 2 1 1 1 1 1 1 2 1 1 2 2 1 2 1 1 2 2 1 2 1 1 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 
## [17461] 2 2 1 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2
## [17497] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 1 1 2 2 1 2 2 2 1 1 1 1 1 2 2 2 2 1
## [17533] 1 1 1 2 1 1 2 2 2 2 2 1 2 1 1 2 1 1 1 2 2 1 2 2 2 1 2 1 1 2 1 2 2 2 1 2 1 1
## [17713] 1 1 1 1 1 1 2 2 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 2 2 2 1 2 2 2 2 1 1 1
## [17821] 2 2 1 2 1 1 2 1 1 2 1 2 1 2 1 1 2 1 1 2 2 2 2 2 2 2 1 2 1 2 1 1 2 1 2 2 2 2
## [17857] 1 2 2 1 1 1 1 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 2 2 2
## [17929] 1 1 1 1 1 1 1 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 2 1 2 1 2 2 2 2 2 1 2 1 2 2
```

```
## [18073] 1 2 1 2 2 1 1 1 1 1 1 2 1 1 2 2 2 1 2 1 1 2 2 2 2 2 1 1 1 2 2 2 2 1 1
## [18109] 2 1 1 2 2 1 2 1 2 1 1 2 1 1 1 2 1 1 1 2 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 1 1
## [18145] 2 1 1 2 2 2 1 2 1 1 2 2 2 1 1 1 1 2 1 2 2 1 2 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1
## [18181] 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 2 2 2 2 1 1 1 2 1 2 1
## [18217] 2 2 2 2 1 2 1 1 2 2 1 2 2 2 2 1 2 2 1 1 1 1 1 2 1 1 2 1 2 2 1 1 1 1 1 1
## [18253] 1 2 1 2 2 2 1 1 2 1 1 2 2 1 2 2 1 1 2 2 1 1 2 2 2 2 1 1 1 2 2 2 2 2 2 2 1
## [18325] 2 2 2 2 1 1 2 1 1 1 2 2 2 2 2 1 2 2 2 1 1 2 1 1 1 1 1 2 1 1 1 1 2 2 2 2 2
## [18361] 2 2 2 2 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 2 2 1 2 1 2 1 2 1 2 1 1
## [18469] 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 2 1
## [18793] 1 2 2 1 2 2 2 2 1 1 2 1 1 2 2 2 2 2 1 2 1 2 1 1 1 2 2 1 1 1 2 2 1 1 1 2 1 1
## [18829] 1 1 1 1 1 1 2 2 2 1 2 2 2 1 1 1 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1
## [18865] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 1 1 1 2 2 1 2 2 2 2 2 1 1 1 1 2 2 1 2 2 2 2 2 1 2 1 1 1 2 2 1 1 2
## [18937] 2 1 1 2 2 2 1 2 1 1 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 2 1 1 1 1 1 1 1 2 1 1
## [19009] 2 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 1 1 2 1 1 2 1 1 2 2 2 2 2 2 2 2
## [19081] 1 2 2 1 2 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 2 1 2 2 2 2 2 2 1 2 1 1 1 1 1 1 1 1
## [19117] 1 2 1 1 2 2 2 2 1 2 2 2 2 1 1 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1 2 1 2 1 2 1 2 1 1 2 1 2 1 1 2 1 2 1 1 2 1 2 1 1 2 1 1 2 1 1 
## [19189] 2 2 2 2 2 2 1 1 1 1 1 1 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 1 1 1 1 1 1 1 2
## [19225] 2 2 2 1 1 1 1 1 1 2 2 1 2 1 2 1 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
## [19261] 2 2 1 1 1 2 1 1 1 1 2 2 1 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 1 1 1 1
## [19369] 2 2 2 2 1 2 2 2 1 1 2 1 1 1 1 1 2 2 2 2 1 1 1 2 1 2 2 2 1 1 1 2 1 2 1 2 1 1 2 1 1 1 2 1 1
## [19441] 2 1 2 2 2 1 2 1 1 1 2 1 2 2 2 1 2 1 1 1 2 2 2 1 2 1 1 1 2 2 2 1 2 2 1 2 1 1 1 2 2
## [19621] 1 1 2 2 2 2 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 2 2 1 2 2 1 1 1 2 1 1 1 1
## [19657] 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 2 2 1 1 2 2 1 1 2 1 1 1 1 1 2 1 2 1 1 1
## [19693] 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 2 1 1 1 1 1 2 1 2 2 2 2
## [19729] 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 1 2 2 2 1 1 2 2 2 1 2 1 2 1 2 1 2 1 2 2 1 1 1 2 2
## [19873] 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1
```

```
## [19981] 1 2 1 1 1 1 2 2 2 1 2 2 2 2 2 2 2 1 1
## Levels: 1 2
# computing accuracy of the model
library(rminer)
print(mmetric(nb_predicts, dtm_train_test_labels, c("ACC")))
## [1] 81.19
pretrained word2vec word embedding
# including the required library
# install.packages("https://cran.r-project.org/src/contrib/Archive/softmaxreg/softmaxreg_1.2.tar.gz",re
library(softmaxreg)
## Attaching package: 'softmaxreg'
## The following object is masked from 'package:e1071':
##
##
      sigmoid
# importing the word2vec pretrained vector into memory
data(word2vec)
# View(word2vec)
dim(word2vec)
## [1] 12853
              21
# function to get word vector for each review
docVectors = function(x) { wordEmbed(x, word2vec, meanVec = TRUE) }
text = read.csv(file='Sentiment Analysis Dataset.csv', header = TRUE)
# applying the docVector function on each of the reviews
# storing the matrix of word vectors as temp
temp=t(sapply(text$SentimentText, docVectors))
# visualizing the word vectors output
# View(temp)
dim(temp)
## [1] 100000
                20
# splitting the dataset into train and test
temp_train=temp[1:N_t,]
temp_test=temp[(N_t+1):N,]
labels_train=as.factor(as.character(text[1:N_t,]$Sentiment))
labels_test=as.factor(as.character(text[(N_t+1):N,]$Sentiment))
# including the random forest library
library(randomForest)
```

```
## randomForest 4.7-1.1
## Type rfNews() to see new features/changes/bug fixes.
# training a model using random forest classifier with training dataset
# observe that we are using 20 trees to create the model
rf_senti_classifier=randomForest(temp_train, labels_train, ntree=20)
print(rf_senti_classifier)
##
## Call:
   randomForest(x = temp_train, y = labels_train, ntree = 20)
##
                  Type of random forest: classification
##
                        Number of trees: 20
## No. of variables tried at each split: 4
##
           OOB estimate of error rate: 39.84%
##
## Confusion matrix:
              2 class.error
        1
## 1 23717 15262 0.3915442
                 0.4050087
## 2 16609 24400
# making predictions on the dataset
rf_predicts<-predict(rf_senti_classifier, temp_test)</pre>
library(rminer)
print(mmetric(rf_predicts, labels_test, c("ACC")))
## [1] 62.45
GloVe word embedding
# including the required library
library(text2vec)
## Attaching package: 'text2vec'
## The following object is masked from 'package:rminer':
##
##
       fit
# reading the dataset
text = read.csv(file='Sentiment Analysis Dataset.csv', header = TRUE)
# subsetting only the review text so as to create Glove word embedding
wiki = as.character(text$SentimentText)
# Create iterator over tokens
tokens = space_tokenizer(wiki)
# Create vocabulary. Terms will be unigrams (simple words).
it = itoken(tokens, progressbar = FALSE)
```

```
vocab = create_vocabulary(it)
# consider a term in the vocabulary if and only if the term has appeared at least three times in the da
vocab = prune_vocabulary(vocab, term_count_min = 3L)
# Use the filtered vocabulary
vectorizer = vocab_vectorizer(vocab)
# use window of 5 for context words and create a term co-occurance matrix
tcm = create_tcm(it, vectorizer, skip_grams_window = 5L)
# create the glove embedding for each in the vocab and
# the dimension of the word embedding should set to 50
# x_max is the maximum number of co-occurrences to use in the weighting
# function
# note that training the word embedding is time consuming - be patient
glove = GlobalVectors$new(rank = 50, x_max = 100)
wv_main = glove$fit_transform(tcm, n_iter = 10, convergence_tol = 0.01)
## INFO [10:46:07.657] epoch 1, loss 0.0501
## INFO [10:46:13.366] epoch 2, loss 0.0318
## INFO [10:46:19.122] epoch 3, loss 0.0267
## INFO [10:46:24.884] epoch 4, loss 0.0240
## INFO [10:46:30.769] epoch 5, loss 0.0222
## INFO [10:46:36.731] epoch 6, loss 0.0209
## INFO [10:46:42.568] epoch 7, loss 0.0199
## INFO [10:46:48.392] epoch 8, loss 0.0191
## INFO [10:46:54.229] epoch 9, loss 0.0185
## INFO [10:47:00.146] epoch 10, loss 0.0179
# Glove model learns two sets of word vectors - main and context.
# both matrices may be added to get the combined word vector
wv_context = glove$components
word_vectors = wv_main + t(wv_context)
# converting the word_vector to a dataframe for visualization
word_vectors=data.frame(word_vectors)
# the word for each embedding is set as row name by default
# using the tibble library rownames_to_column function, the rownames is copied as first column of the d
# we also name the first column of the dataframe as words
library(tibble)
word_vectors=rownames_to_column(word_vectors, var = "words")
# View(word vectors)
library(softmaxreg)
docVectors = function(x) { wordEmbed(x, word_vectors, meanVec = TRUE) }
# applying the function docVectors function on the entire reviews dataset
# this will result in word embedding representation of the entire reviews
# dataset
temp=t(sapply(text$SentimentText, docVectors))
# View(temp)
# splitting the dataset into train and test portions
temp_train=temp[1:N_t,]
temp_test=temp[(N_t+1):N,]
labels train=as.factor(as.character(text[1:N t,]$Sentiment))
labels_test=as.factor(as.character(text[(N_t+1):N,]$Sentiment))
```

```
rf_senti_classifier=randomForest(temp_train, labels_train, ntree=20)
print(rf_senti_classifier)
##
## Call:
## randomForest(x = temp_train, y = labels_train, ntree = 20)
                  Type of random forest: classification
##
                        Number of trees: 20
## No. of variables tried at each split: 7
           OOB estimate of error rate: 32.13%
##
## Confusion matrix:
               2 class.error
         1
## 1 26652 12328 0.3162648
## 2 13376 27634
                 0.3261644
# predicting labels using the randomforest model created
rf_predicts<-predict(rf_senti_classifier, temp_test)</pre>
# estimating the accuracy from the predictions
library(rminer)
print(mmetric(rf_predicts, labels_test, c("ACC")))
## [1] 71.755
fastText word embedding
# loading the required libary
library(fastTextR)
# reading the input reviews file
# recollect that fastText needs the file in a specific format and we created one compatiable file in
# "Understanding the Amazon Reviews Dataset" section of this chapter
text = readLines("Sentiment Analysis Dataset_ft.txt")
# Viewing the text vector for conformation
# View(text)
# dividing the reviews into training and test
temp_train=text[1:N_t]
temp_test=text[(N_t+1):N]
# Viewing the train datasets for confirmation
# View(temp train)
# creating txt file for train and test dataset
# the fasttext function expects files to be passed for training and testing
fileConn<-file("train.ft.txt")</pre>
writeLines(temp_train, fileConn)
close(fileConn)
fileConn<-file("test.ft.txt")</pre>
writeLines(temp_test, fileConn)
close(fileConn)
```

using randomforest to build a model on train data library(randomForest)

```
# creating a test file with no labels
# recollect the original test dataset has labels in it
# as the dataset is just a subset obtained from full dataset
temp_test_nolabel<- gsub("__label__1", "", temp_test, perl=TRUE)</pre>
temp_test_nolabel<- gsub("__label__2", "", temp_test_nolabel, perl=TRUE)</pre>
# View(temp_test_nolabel)
fileConn<-file("test_nolabel.ft.txt")</pre>
writeLines(temp_test_nolabel, fileConn)
close(fileConn)
# training a supervised classification model with training dataset file
model<-ft_train("train.ft.txt", method = "supervised", control = ft_control(nthreads = 3L, seed = 1))</pre>
# Obtain all the words from a previously trained model=
words<-ft_words(model)</pre>
# viewing the words for confirmation. These are the set of words present in our training data
# View(words)
# Obtain word vectors from a previously trained model.
word_vec<-ft_word_vectors(model, words)</pre>
# Viewing the word vectors for each word in our training dataset
# observe that the word embedding dimension is 5
# View(word_vec)
# predicting the labels for the reviews in the no labels test dataset
# qetting the predictions into a dataframe so as to compute performance measurement
ft_preds<-ft_predict(model, newdata = temp_test_nolabel)</pre>
# reading the test file to extract the actual labels
reviewstestfile<- readLines("test.ft.txt")</pre>
# extracting just the labels frm each line
library(stringi)
actlabels<-stri_extract_first(reviewstestfile, regex="\\w+")</pre>
# converting the actual labels and predicted labels into factors
actlabels<-as.factor(as.character(actlabels))</pre>
ft_preds<-as.factor(as.character(ft_preds$label))</pre>
# getting the estimate of the accuracy
library(rminer)
print(mmetric(actlabels, ft_preds, c("ACC")))
```

[1] 86.495

Drug Data

```
N_Drug = 146942
N_t_Drug = round(0.8*N_Drug)
reviews_text_Drug<-readLines("Drug Train.csv", n = N_Drug)
reviews_text_Drug<-data.frame(reviews_text_Drug)</pre>
```

```
library(tidyr)
reviews_text_Drug<-separate(data = reviews_text_Drug, col = reviews_text_Drug, into = c("Sentiment", "S</pre>
```

```
reviews_text_Drug<-reviews_text_Drug[-1,]</pre>
N_Drug = N_Drug - 1
# Retaining only alphanumeric values in the sentiment column
reviews_text_Drug$Sentiment<-gsub("[^[:alnum:]]","",reviews_text_Drug$Sentiment)
# Retaining only alphanumeric values in the sentiment text
reviews_text_Drug$SentimentText<-gsub("[^[:alnum:]]"," ",reviews_text_Drug$SentimentText)
# Replacing multiple spaces in the text with single space
reviews_text_Drug$SentimentText<-gsub("(?<=[\\s])\\s*|^\\s+$", "", reviews_text_Drug$SentimentText
# Writing the output to a file that can be consumed in other projects
write.table(reviews_text_Drug,file = "Sentiment Analysis Dataset_Drug.csv",row.names = F,col.names = T,
# reading the first 1000 reviews from the dataset
reviews_text_Drug<-readLines("Drug Train.csv", n = N_Drug)</pre>
# basic EDA to confirm that the data is read correctly
print(class(reviews_text_Drug))
## [1] "character"
print(length(reviews_text_Drug))
## [1] 146941
print(head(reviews_text_Drug,2))
## [1] "\"Sentiment\",\"SentimentText\""
## [2] "\"2\",\"Valsartan Left Ventricular Dysfunction It has no side effect I take it in combination o
# replacing the positive sentiment value 2 with __label__2
reviews_text_Drug<-gsub("\\\",",","__label__2 ",reviews_text_Drug)</pre>
# replacing the negative sentiment value 1 with __label__1
reviews_text_Drug<-gsub("\\\",","_label_1 ",reviews_text_Drug)</pre>
# removing the unnecessary \" characters
reviews_text_Drug<-gsub("\\\""," ",reviews_text_Drug)</pre>
# replacing multiple spaces in the text with single space
reviews_text_Drug<-gsub("(?<=[\\s])\\s*|^\\s+\\, "", reviews_text_Drug, perl=TRUE)
# Basic EDA post the required processing to confirm input is as desired
print("EDA POST PROCESSING")
## [1] "EDA POST PROCESSING"
print(class(reviews_text_Drug))
## [1] "character"
print(length(reviews_text_Drug))
```

[1] 146941

```
print(head(reviews_text_Drug,2))
## [1] "Sentiment , SentimentText"
## [2] "__label__2 Valsartan Left Ventricular Dysfunction It has no side effect I take it in combination
# writing the revamped file to the directory so we could use it with
# fastText sentiment analyzer project
fileConn<-file("Sentiment Analysis Dataset_ft_Drug.txt")</pre>
writeLines(reviews_text_Drug, fileConn)
close(fileConn)
BoW approach
# including the required libraries
library(SnowballC)
library(tm)
# setting the working directory where the text reviews dataset is located
# reading the transformed file as a dataframe
text_Drug <- read.table(file='Sentiment Analysis Dataset_Drug.csv', sep=',', header = TRUE)
# checking the dataframe to confirm everything is in tact
print(dim(text_Drug))
## [1] 146941
# View(text)
# transforming the text into volatile corpus
train_corp_Drug = VCorpus(VectorSource(text_Drug$SentimentText))
print(train_corp_Drug)
## <<VCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 146941
# creating document term matrix
dtm_train_Drug <- DocumentTermMatrix(train_corp_Drug, control = list( tolower = TRUE, removeNumbers = TR
# Basic EDA on dtm
inspect(dtm_train_Drug)
## <<DocumentTermMatrix (documents: 146941, terms: 61384)>>
## Non-/sparse entries: 5387849/9014438495
## Sparsity
                     : 100%
## Maximal term length: 95
## Weighting
                  : term frequency (tf)
## Sample
##
## Docs
          day effect get month pain start take week work year
    102079 6
                0 2
                              0
                                   7
                                         0
                                              2
```

```
127353
##
                     1
                                    1
                                                     1
                                                          1
                                                               5
##
     128846 10
                     0
                         4
                               1
                                    9
                                           2
                                                2
                                                     3
                                                          0
                                                               3
                         5
                               2
                                                          2
                                                               3
##
     136794 5
                     3
                                   11
                                           1
##
     139424 0
                     2
                         2
                               4
                                           4
                                                          2
                                    3
                                                1
                                                     0
                                                               1
                         2
                                    7
                                                2
##
     20066
              6
                     0
                               0
                                           0
                                                     0
                                                          4
                                                               2
##
     39445
             7
                     7
                         4
                               1
                                    3
                                          1
                                                6
                                                     4
                                                          4
                                                               0
##
     63207
             7
                     7
                         4
                                    3
                                                6
                                                          4
                                                               0
##
     69005
              7
                         3
                               2
                                          4
                                                9
                                                          2
                     5
                                    0
                                                     6
                                                               1
##
     85059
                     2
                                    3
                                           4
# Removing sparse terms
dtm_train_Drug = removeSparseTerms(dtm_train_Drug, 0.99)
inspect(dtm_train_Drug)
## <<DocumentTermMatrix (documents: 146941, terms: 633)>>
## Non-/sparse entries: 3988263/89025390
## Sparsity
                      : 96%
## Maximal term length: 14
## Weighting
                      : term frequency (tf)
## Sample
##
           Terms
            day effect get month pain start take week work year
## Docs
##
     127353
             9
                     1
                         1
                               1
                                    1
                                           3
                                                     1
                                                          1
                         4
                                    9
                                           2
                                                2
                                                               3
##
     128846 10
                     0
                               1
                                                     3
                                                          0
##
     136794
            5
                     3
                         5
                               2
                                                6
                                                     2
                                                          2
                                                               3
                                   11
                                           1
##
     139424 0
                     2
                        2
                                    3
                                                1
              7
                     7
##
     39445
                         4
                               1
                                    3
                                           1
                                                6
                                                     4
                                                          4
                                                               0
                                    2
##
     44457
              0
                     6
                        4
                               0
                                           0
                                               10
                                                     0
                                                          5
                                                               0
##
     63207
             7
                     7
                         4
                               1
                                    3
                                          1
                                                6
                                                     4
                                                          4
                                                               0
                         3
##
     69005
           7
                     5
                               2
                                  0
                                           4
                                                9
                                                          2
                         2
##
     85059
              0
                     2
                               4
                                    3
                                           4
                                                          2
                                                1
                                                     0
                                                               1
##
     95250
                     4
                                   10
                                           4
# splitting the train and test DTM
dtm train train Drug <- dtm train Drug[1:N t Drug, ]
dtm_train_test_Drug <- dtm_train_Drug[(N_t_Drug+1):N_Drug, ]</pre>
dim(dtm_train_Drug)
## [1] 146941
                 633
dtm_train_train_Drug_labels <- as.factor(as.character(text_Drug[1:N_t_Drug, ]$Sentiment))</pre>
dtm_train_test_Drug_labels <- as.factor(as.character(text_Drug[(N_t_Drug+1):N_Drug, ]$Sentiment))</pre>
cellconvert<- function(x) { x <- ifelse(x > 0, "Y", "N") }
# applying the function to rows in training and test datasets
dtm_train_train_Drug <- apply(dtm_train_train_Drug, MARGIN = 2,cellconvert)</pre>
dtm_train_test_Drug <- apply(dtm_train_test_Drug, MARGIN = 2,cellconvert)</pre>
# inspecting the train dtm to confirm all is in tact
# View(dtm_train_train)
```

training model

##

##

##

```
# training the naive bayes classifier on the training dtm
library(e1071)
nb_senti_classifier_Drug=naiveBayes(dtm_train_train_Drug,dtm_train_train_Drug_labels)
# printing the summary of the model created
summary(nb_senti_classifier_Drug)
##
     Length Class Mode
## apriori
      2
        table numeric
## tables
     633
        -none- list
## levels
      2
        -none- character
## isnumeric 633
        -none- logical
## call
        -none- call
# making predictions on the test data dtm
nb_predicts_Drug <- predict(nb_senti_classifier_Drug, dtm_train_test_Drug,type="class")</pre>
# printing the predictions from the model
print(nb_predicts_Drug)
  ##
  [37] 2 2 2 2 2 2 1 2 2 2 1 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2
##
  [73] \ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2
##
##
  ##
  ##
##
  ##
  ##
  ##
  ##
  ##
  [397] 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 1 1 1 2 1 2 2 1 2
##
  [433] 1 2 2 2 2 1 2 1 2 1 1 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2
##
  ##
  ##
##
  [577] 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 1 1 1 1 2 2 1 1 2 1 2 2 1 2 2 1
  ##
  ##
##
  ##
##
  [757] 2 1 2 2 1 1 1 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1
  ##
##
  ##
  ##
##
  ##
```

[1225] 2 1 2 1 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 1 2 2 1 1 1 2 2 2 2 1 1 ## ## ## ## ## [1441] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 1 2 2 1 1 2 2 1 2 2 1 1 2 2 1 2 2 1 1 2 2 $[1477] \ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1$ ## ## ## ## $[1657] \ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2$ ## ## ## ## ## ## ## ## ## ## ## ## ## ## ## ## [2197] 1 2 1 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1 1 1 1 1 ## ## ## ## [2377] 2 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 ## ## ## ## ## ## ## ## ## ## ## ## ## [2809] 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 1 2 2 ## ## $\begin{smallmatrix} 2917 \end{smallmatrix}] \ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2$ ## ## ## ## ##

[5041] 1 2 2 2 2 2 2 2 1 1 1 2 1 1 2 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 1 1 1 1 1 1 2 $[5077] \ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2$ ## ## ## ## ## [5257] 2 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 2 1 2 [5293] 2 2 1 1 2 2 2 2 2 1 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 1 1 1 1 1 2 1 1 ## ## [5329] 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 1 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 ## ## ## [5473] 2 1 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 1 1 2 2 2 1 1 2 2 2 1 1 2 2 1 1 2 ## ## $[5545] \ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2$ ## ## ## ## ## ## ## ## ## ## ## [5941] 1 2 2 2 2 1 1 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 ## $[6013] \ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$ ## $[6049] \ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1$ ## ## ## ## ## ## ## [6301] 1 2 2 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 ## ## ## $[6409] \ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 1$ ## ## ## ## ## ## ## ## ## ## ## $[6805] \ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$ ## ## ## ## ##

[7237] 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 ## ## $[7309] \ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2$ ## [7381] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 ## ## [7417] 1 2 2 1 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 [7453] 2 2 2 1 2 2 1 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2 1 2 1 1 1 2 1 2 2 2 1 2 ## ## ## $[7561] \ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 1\ 2$ ## ## ## ## ## $[7705] \ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2$ ## ## ## ## ## ## $[7921] \ 2\ 2\ 1\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1$ ## ## ## ## ## ## $[8173] \ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1$ ## ## [8209] 2 1 1 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 1 2 1 2 2 2 2 2 ## [8245] 2 2 2 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 ## ## ## ## ## ## [8497] 1 2 1 1 2 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 ## $[8533] \ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2$ ## ## ## $[8641] \ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$ ## ## ## $[8713] \ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2$ ## ## ## ##

$[9037] \ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2$ ## ## ## ## $[9181] \ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1$ ## ## ## ## ## ## ## $[9433] \ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2$ ## ## ## ## ## ## ## $[9685] \ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 1$ ## ## ## ## ## ## $[9901] \ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 1$ ## ## $[10045] \ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 1\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 2\ 2\ 2\ 2\ 1$ ## [10081] 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 2 2 1 1 2 2 2 1 1 2 2 2 2 ## [10117] 2 2 2 2 1 1 1 1 1 1 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 1 ## [10153] 2 2 1 2 1 2 2 1 2 2 2 2 2 1 2 2 2 1 1 1 2 2 1 1 2 1 1 1 1 2 2 2 2 2 ## [10189] 2 1 2 1 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 1 2 1 1 2 1 1 2 2 2 ## [10369] 2 1 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 2 2 2 1 1 2 2 ## [10405] 2 2 2 1 2 2 2 2 2 1 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 1 2 2 2 ## [10513] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2 2 2 1 2 1 1 2 ## [10657] 2 2 2 2 2 1 1 1 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1 2 ## [10765] 1 2 1 2 2 2 1 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2

```
## [11197] 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 2 1 2 1 2 2 2 2 1 1 1 1 1 2
## [11269] 2 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 1 1 2 2 1 1 1 2 1
## [11521] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 1 1 2 2 2 2 1 1 1 2 2 2 2
## [11593] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 2 2 2 1 2 1 1 2 1 1 2 1 1 2
## [11665] 2 1 2 1 2 2 2 1 2 1 1 2 2 2 2 1 2 1 1 1 2 2 2 2 1 1 1 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2
## [11845] 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 1 1 1 1 2 1 1 2 2 2 2 2 1 2 2 2 2 2
## [12097] 1 1 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 1 2 2 1 2
## [12313] 2 2 1 2 2 1 2 1 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 1 1
## [12637] 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1
## [12709] 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2
```

```
## [13033] 2 2 2 2 2 1 2 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 2 2 2 2 1 2 2
## [13069] 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 1 1
## [13285] 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2
## [13429] 1 2 2 2 1 2 1 2 2 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1
## [13465] 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 2 1 2 2
## [13645] 2 1 1 1 2 2 2 2 2 2 1 2 1 1 2 1 2 2 1 2 1 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2
## [13789] 2 2 2 2 1 1 1 2 2 1 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 
## [13861] 2 2 1 1 1 1 1 2 1 1 2 1 1 1 2 1 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 1 2 2 1
## [14005] 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 1 2 2 2 2 2 1 1 2 1 2 2
## [14041] 2 1 2 2 2 2 2 2 1 2 2 1 1 2 1 2 1 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1 2 1 2 1 2 1 2
## [14113] 2 2 2 2 1 1 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2
## [14221] 2 1 2 1 2 2 1 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2
## [14329] 2 1 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 1
## [14365] 2 2 2 2 2 2 1 1 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 1
## [14401] 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2
## [14437] 1 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1
## [14545] 2 2 2 1 2 2 1 1 2 2 2 1 2 1 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2
## [14617] 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 1 1 1 2 2 1 2 2 2 2 2 2 1 2
## [14653] 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2
## [14725] 2 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 1 1 2 2 1 2 2 1 2 2 2 2 1 2 1 1
```

```
## [15049] 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 1 2
## [15373] 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 1 2 2 1 1 2 2 2
## [15697] 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 1 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 1
## [15769] 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 1 2 2 2 1 1 2 2 2 1 1 2 2 2 1 2 2 2 2 1 2
## [15805] 2 1 1 1 2 2 2 2 2 1 2 1 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 1
## [15841] 2 2 2 2 2 1 2 2 1 2 2 1 1 2 1 1 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 2 2 1
## [15985] 1 2 2 2 2 1 2 2 2 2 2 1 2 1 1 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2
## [16093] 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2 1 1 2 2 2 2
## [16129] 2 2 1 2 1 2 2 2 1 1 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 2 2 2 2 1 1 2 1 2 2 2
## [16273] 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 1 2 2
## [16345] 2 1 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 1 2 1 2 2 1 2 1 2 2 2 2 2 2 1 2
## [16417] 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 1 1 2 2 1 1 2 2 2
```

```
## [16705] 1 2 2 1 2 1 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1 2 1 2 2 2 2 2 2 1 1 1 2
## [16741] 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 2 1 2 2 2 1 1 1 2 2 2 2 2
## [16813] 2 1 2 2 1 2 2 1 1 1 1 1 2 1 2 2 2 1 1 2 1 1 2 1 2 2 2 2 2 2 2 1 2 1 1 2 2 2
## [16885] 2 1 2 1 1 2 1 2 1 1 2 1 1 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 1 2 2 1 2 2
## [17065] 2 1 1 2 1 2 2 1 1 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 1 2 1 1 1 1 2 2 2 2 2 2 2 2
## [17281] 1 2 1 2 2 1 2 2 1 1 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 2 1 1 1 2 1 2 1 2 2 2 2 2 2 1 2 2
## [17317] 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 1 1 2 2 1 1 2 2 2 1 1 2
## [17389] 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2
## [17461] 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 1 2 2 1 2 2 1 1 2 2 2 2 2 1 2 1 1 2 1 2 2 2 1 1 2
## [17713] 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 1 2 2 2 1 1 2 2 2 1 2 2 1 2 2
## [18001] 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 2 2 2 1 2 1 2 1 2 2 2 1 1 2 2 2 1
## [18145] 1 2 2 1 2 2 1 1 2 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 1
## [18289] 2 2 1 2 1 1 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2
## [18325] 2 2 1 2 2 1 1 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 1 2 1 1 1 2 1 2 2 2 2 2 2 2 2
## [18469] 2 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2
## [18577] 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2 1 1 2 2 1 2 2 2 2 1 1 2 1 2 1 2 1
```

```
## [18721] 1 1 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2
## [18793] 2 2 1 2 2 2 1 2 2 2 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1 2 1 1 1 2 1 1 2 1 2 2
## [18829] 2 1 1 2 2 1 2 1 2 2 1 2 2 2 1 1 1 2 1 2 2 1 2 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2
## [18901] 1 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 1 1 2 2 1 1 2 2 2 1 2 2
## [18973] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2
## [19117] 1 2 1 2 2 2 1 2 1 1 1 2 2 2 2 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 2 1
## [19153] 1 2 1 2 2 2 1 2 2 2 1 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 1 2 1
## [19189] 1 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1
## [19225] 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 1 1 2 2 1 2 1 2
## [19333] 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 1 1 2 2 2 1 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1
## [19729] 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1
## [20017] 2 2 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 2 2 2 2 1 1 1 1 2 2 2 2 2 2 1 2 1 1 1 2 2 2
## [20269] 2 2 1 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 2 1 1 1 2 1 2 2 2 1 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 1 1
## [20485] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1
```

```
## [20665] 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 1 1 2 2 1 1 2 2 2 2 1 2 2 1 1 2 2 1 2
## [20917] 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2
## [20953] 2 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 1 2 2 2
## [21061] 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1 1 2 1 2 1 2 1 2 1 2 2 1 2 2 2 2
## [21169] 2 2 1 1 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 1 2 2 2 1 1 1 2 2
## [21205] 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2
## [21241] 2 1 1 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 1 2 1
## [21277] 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 1
## [21349] 2 1 2 1 1 2 2 1 2 2 2 1 1 1 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 5
## [21385] 1 2 2 1 2 2 1 2 1 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 1
## [21601] 2 2 2 2 1 2 2 1 2 2 2 2 2 1 2 1 2 1 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 1
## [21637] 1 2 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1
## [21745] 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2
## [21889] 2 1 2 2 1 2 1 2 1 1 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2
## [21997] 1 2 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 1 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1
## [22069] 1 2 2 2 2 1 1 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 1
## [22105] 1 2 2 2 1 1 2 1 1 2 2 2 1 2 1 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 2 1 2 1 2 2 2 1 2 1 2 2 2 1
## [22141] 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 2
## [22285] 1 2 2 2 2 2 2 2 2 1 2 1 1 1 2 1 2 1 1 2 1 2 1 2 2 2 2 2 2 1 2 1 1 1 2 2
## [22357] 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 2 2 1 1
## [22465] 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 1 1 2 2 2 1 1 2 2 1 1 1 1 1 2 2 1 2 2 1
## [22501] 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 2 1 2 1 2
```

```
## [22609] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1
## [22933] 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2
## [23005] 2 2 1 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 1 2 1 2 1 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2
## [23077] 2 2 2 2 1 1 2 1 2 2 2 1 2 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2
## [23113] 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 2 1
## [23401] 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2
## [23437] 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2
## [23473] 2 2 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
## [23725] 2 2 2 2 1 1 2 1 2 2 2 1 2 2 2 1 2 2 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1 2 1
## [23797] 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 1 1 1 2 1 2 1 1 1 1 2 2 1 1 2 1 2
## [23905] 1 1 2 1 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 1 1 2 2 1 2 1 1 2 2 1 1 1
## [23977] 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 1 2 1 2 1 2 1 2 1 2 2 2 2 2 1 2 1
## [24049] 2 1 1 2 1 2 1 1 2 2 2 2 2 1 1 1 2 2 1 2 1 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2 1 2 1 2 2 2 1 2 1
## [24085] 2 1 2 2 1 1 2 1 1 1 2 2 2 2 2 2 1 1 2 2 1 2 2 1 2 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 
## [24121] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1
## [24265] 2 2 2 1 2 2 2 1 2 1 1 1 1 2 2 2 2 2 1 2 1 1 2 1 1 2 1 2 2 1 2 2 2 2 2 2
## [24445] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1
```

```
## [24769] 2 2 2 2 2 2 1 2 1 2 1 1 1 1 2 1 2 2 1 1 1 1 2 2 2 1 1 2 1 2 2 1 2 1 2 1 2 1 2
## [24913] 2 2 2 1 2 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2
## [25021] 2 1 2 1 2 1 1 2 1 1 2 1 2 1 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 1
## [25057] 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 1 2 2 1 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 1 1 2 1 2 1 2 1 2
## [25093] 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 1 2
## [25165] 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2
## [25273] 1 1 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1
## [25417] 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 1 2 2 2 2 2 2 1 2 2 2 1 1 1 1
## [25561] 2 1 2 2 2 2 2 1 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2
## [25705] 2 1 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 1 2 2 2 2 2 2 2 1
## [25813] 1 2 2 1 2 2 2 2 2 1 2 2 1 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 1 2 2 2 1 1 1 1 1 2 2 2
## [25849] 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 1 1 2 1 1 1 2 2 2 2 2
## [25993] 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1 2 2 1 2 1 1
## [26173] 2 1 2 2 1 2 1 1 2 1 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2
## [26281] 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2
## [26317] 1 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 1 1
```

```
## [26497] 2 2 2 2 1 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 2
## [26569] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 1 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2
## [26677] 2 1 2 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2
## [26821] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 1 2 1 2 1 1 2 2 2 1 2 2 1 1
## [27109] 2 1 2 1 2 1 2 2 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1
## [27217] 2 2 2 2 2 1 1 2 1 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2
## [27325] 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 1 2
## [27433] 2 1 2 2 2 1 1 2 2 2 1 2 1 2 1 2 1 1 1 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1
## [27541] 2 1 1 1 2 1 2 2 1 2 2 1 2 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1
## [27685] 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2
## [27937] 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 1 2 2 1 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2
## [28045] 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 2 1 2
```

```
## [28405] 2 2 1 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 1 2 1 2 2 1 2 1 2 2 2 2 1 2 2
## [28657] 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 1
## [28837] 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 1 2 1 1 2 1 1
## [28873] 2 1 1 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2
## [28945] 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 1 1 2 2 1 2 1 2 1 2 2 1 1 1 2 2 1 1 2 2 1 2
## [28981] 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 1 1 1 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2
## [29017] 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
## [29053] 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 1 1 1 1 2 1 2 2 1 2 1 1
## [29089] 2 1 1 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2 1 1 2 1 2 1 2 2 1 2 1 2 2 2 2 2 1 1
## [29161] 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 1 2 1 2 1 2 1 1 1 2 2 1 2 2 1 2 2 2
## [29377] 1 1 2 2 2 2 2 1 1 2 2
## Levels: 1 2
# computing accuracy of the model
library(rminer)
print(mmetric(nb_predicts_Drug, dtm_train_test_Drug_labels, c("ACC")))
```

[1] 79.11662

pretrained word2vec word embedding

```
# including the required library
# install.packages("https://cran.r-project.org/src/contrib/Archive/softmaxreg/softmaxreg_1.2.tar.gz",re
library(softmaxreg)
# importing the word2vec pretrained vector into memory
data(word2vec)
# View(word2vec)
dim(word2vec)
```

[1] 12853 21

```
# function to get word vector for each review
docVectors = function(x) { wordEmbed(x, word2vec, meanVec = TRUE) }
text Drug = read.csv(file='Sentiment Analysis Dataset Drug.csv', header = TRUE)
# applying the docVector function on each of the reviews
# storing the matrix of word vectors as temp
temp_Drug=t(sapply(text_Drug$SentimentText, docVectors))
# visualizing the word vectors output
# View(temp)
dim(temp_Drug)
## [1] 146941
                  20
# splitting the dataset into train and test
temp_train_Drug=temp_Drug[1:N_t_Drug,]
temp_test_Drug=temp_Drug[(N_t_Drug+1):N_Drug,]
labels_train_Drug=as.factor(as.character(text_Drug[1:N_t_Drug,]$Sentiment))
labels test Drug=as.factor(as.character(text Drug[(N t Drug+1):N Drug,]$Sentiment))
# including the random forest library
library(randomForest)
# training a model using random forest classifier with training dataset
# observe that we are using 20 trees to create the model
rf_senti_classifier_Drug=randomForest(temp_train_Drug, labels_train_Drug, ntree=20)
print(rf_senti_classifier_Drug)
##
## Call:
## randomForest(x = temp_train_Drug, y = labels_train_Drug, ntree = 20)
                  Type of random forest: classification
##
##
                        Number of trees: 20
## No. of variables tried at each split: 4
##
           OOB estimate of error rate: 19.88%
##
## Confusion matrix:
              2 class.error
## 1 16142 15993 0.49768166
## 2 7373 78033 0.08632883
# making predictions on the dataset
rf_predicts_Drug<-predict(rf_senti_classifier_Drug, temp_test_Drug)
library(rminer)
print(mmetric(rf_predicts_Drug, labels_test_Drug, c("ACC")))
## [1] 83.11839
GloVe word embedding
```

```
# including the required library
library(text2vec)
# reading the dataset
text_Drug = read.csv(file='Sentiment Analysis Dataset_Drug.csv', header = TRUE)
```

```
# subsetting only the review text so as to create Glove word embedding
wiki_Drug = as.character(text_Drug$SentimentText)
# Create iterator over tokens
tokens_Drug = space_tokenizer(wiki_Drug)
# Create vocabulary. Terms will be unigrams (simple words).
it_Drug = itoken(tokens_Drug, progressbar = FALSE)
vocab_Drug = create_vocabulary(it_Drug)
# consider a term in the vocabulary if and only if the term has appeared at least three times in the da
vocab_Drug = prune_vocabulary(vocab_Drug, term_count_min = 3L)
# Use the filtered vocabulary
vectorizer_Drug = vocab_vectorizer(vocab_Drug)
# use window of 5 for context words and create a term co-occurance matrix
tcm_Drug = create_tcm(it_Drug, vectorizer_Drug, skip_grams_window = 5L)
# create the glove embedding for each in the vocab and
# the dimension of the word embedding should set to 50
# x_max is the maximum number of co-occurrences to use in the weighting
# function
# note that training the word embedding is time consuming - be patient
glove = GlobalVectors$new(rank = 50, x_max = 100)
wv_main_Drug = glove$fit_transform(tcm_Drug, n_iter = 10, convergence_tol = 0.01)
## INFO [10:59:21.808] epoch 1, loss 0.0859
## INFO [10:59:25.239] epoch 2, loss 0.0541
## INFO [10:59:28.638] epoch 3, loss 0.0442
## INFO [10:59:32.137] epoch 4, loss 0.0390
## INFO [10:59:35.568] epoch 5, loss 0.0358
## INFO [10:59:39.045] epoch 6, loss 0.0335
## INFO [10:59:42.449] epoch 7, loss 0.0318
## INFO [10:59:45.933] epoch 8, loss 0.0305
## INFO [10:59:49.589] epoch 9, loss 0.0294
## INFO [10:59:52.981] epoch 10, loss 0.0285
# Glove model learns two sets of word vectors - main and context.
# both matrices may be added to get the combined word vector
wv context = glove$components
word_vectors_Drug = wv_main_Drug + t(wv_context)
# converting the word_vector to a dataframe for visualization
word_vectors_Drug=data.frame(word_vectors_Drug)
# the word for each embedding is set as row name by default
# using the tibble library rownames_to_column function, the rownames is copied as first column of the d
# we also name the first column of the dataframe as words
library(tibble)
word_vectors_Drug=rownames_to_column(word_vectors_Drug, var = "words")
# View(word_vectors)
library(softmaxreg)
docVectors_Drug = function(x) { wordEmbed(x, word_vectors_Drug, meanVec = TRUE) }
# applying the function docVectors function on the entire reviews dataset
# this will result in word embedding representation of the entire reviews
# dataset
temp_Drug=t(sapply(text_Drug$SentimentText, docVectors_Drug))
# View(temp)
```

```
temp_train_Drug=temp_Drug[1:N_t_Drug,]
temp test Drug=temp Drug[(N t Drug+1):N Drug,]
labels_train_Drug=as.factor(as.character(text_Drug[1:N_t_Drug,]$Sentiment))
labels test Drug=as.factor(as.character(text Drug[(N t Drug+1):N Drug,]$Sentiment))
# using randomforest to build a model on train data library(randomForest)
rf_senti_classifier_Drug=randomForest(temp_train_Drug, labels_train_Drug,ntree=20)
print(rf_senti_classifier_Drug)
##
## Call:
## randomForest(x = temp_train_Drug, y = labels_train_Drug, ntree = 20)
                  Type of random forest: classification
##
                        Number of trees: 20
## No. of variables tried at each split: 7
          OOB estimate of error rate: 20.65%
##
## Confusion matrix:
             2 class.error
        1
## 1 15707 16430 0.51124872
## 2 7847 77557 0.09188094
# predicting labels using the randomforest model created
rf_predicts_Drug<-predict(rf_senti_classifier_Drug, temp_test_Drug)
# estimating the accuracy from the predictions
library(rminer)
print(mmetric(rf_predicts_Drug, labels_test_Drug, c("ACC")))
## [1] 82.7815
fastText word embedding
# loading the required libary
library(fastTextR)
# reading the input reviews file
# recollect that fastText needs the file in a specific format and we created one compatiable file in
# "Understanding the Amazon Reviews Dataset" section of this chapter
text_Drug = readLines("Sentiment Analysis Dataset_ft_Drug.txt")
# Viewing the text vector for conformation
# View(text)
# dividing the reviews into training and test
temp_train_Drug=text_Drug[1:N_t_Drug]
temp test Drug=text Drug[(N t Drug+1):N Drug]
# Viewing the train datasets for confirmation
# View(temp_train)
```

splitting the dataset into train and test portions

the fasttext function expects files to be passed for training and testing

creating txt file for train and test dataset

fileConn<-file("train_Drug.ft.txt")</pre>

```
writeLines(temp_train_Drug, fileConn)
close(fileConn)
fileConn<-file("test_Drug.ft.txt")</pre>
writeLines(temp_test_Drug, fileConn)
close(fileConn)
# creating a test file with no labels
# recollect the original test dataset has labels in it
# as the dataset is just a subset obtained from full dataset
temp_test_Drug_nolabel<- gsub("__label__1", "", temp_test_Drug, perl=TRUE)</pre>
temp_test_Drug_nolabel<- gsub("__label__2", "", temp_test_Drug_nolabel, perl=TRUE)</pre>
# View(temp_test_nolabel)
fileConn<-file("test_Drug_nolabel.ft.txt")</pre>
writeLines(temp_test_Drug_nolabel, fileConn)
close(fileConn)
# training a supervised classification model with training dataset file
model_Drug<-ft_train("train_Drug.ft.txt", method = "supervised", control = ft_control(nthreads = 3L, se</pre>
# Obtain all the words from a previously trained model=
words_Drug<-ft_words(model_Drug)</pre>
# viewing the words for confirmation. These are the set of words present in our training data
# View(words)
# Obtain word vectors from a previously trained model.
word_vec_Drug<-ft_word_vectors(model_Drug, words_Drug)</pre>
# Viewing the word vectors for each word in our training dataset
# observe that the word embedding dimension is 5
# View(word_vec)
# predicting the labels for the reviews in the no labels test dataset
# qetting the predictions into a dataframe so as to compute performance measurement
ft_preds_Drug<-ft_predict(model_Drug, newdata = temp_test_Drug_nolabel)</pre>
# reading the test file to extract the actual labels
reviewstestfile_Drug<- readLines("test_Drug.ft.txt")</pre>
# extracting just the labels frm each line
library(stringi)
actlabels_Drug<-stri_extract_first(reviewstestfile_Drug, regex="\\w+")</pre>
# converting the actual labels and predicted labels into factors
actlabels_Drug<-as.factor(as.character(actlabels_Drug))</pre>
ft_preds_Drug<-as.factor(as.character(ft_preds_Drug$label))</pre>
# getting the estimate of the accuracy
library(rminer)
print(mmetric(actlabels_Drug, ft_preds_Drug, c("ACC")))
```

[1] 86.26944