
title: "R Notebook for Naive Bayes Claasifier" output: html_notebook

```
#Using Bags of Words and Naive Bayes for SMS spam classification
```

```
sms_raw = read.csv("https://raw.githubusercontent.com/bhagi8289/datasets/master/sms_spam.csv", stringsAsFactors = FALSE)
str(sms_raw)
```

```
## 'data.frame': 5559 obs. of 2 variables:
## $ type: chr "ham" "ham" "ham" "spam" ...
## $ text: chr "Hope you are having a good week. Just checking in" "K..give back my thanks." "Am also"
```

```
sms_raw$type <- factor(sms_raw$type)
str(sms_raw)
```

```
## 'data.frame': 5559 obs. of 2 variables:
## $ type: Factor w/ 2 levels "ham","spam": 1 1 1 2 2 1 1 1 2 1 ...
## $ text: chr "Hope you are having a good week. Just checking in" "K..give back my thanks." "Am also"
```

```
table(sms_raw$type)
```

```
##
## ham spam
## 4812 747
```

```
# Data Cleaning for Text Data
```

```
# Use text-mining package tm
```

```
# Create corpus - a collection of text documents i.e. collection of SMS messages
```

```
library(tm)
```

```
## Loading required package: NLP
```

```
sms_corpus <- VCorpus(VectorSource(sms_raw$text))
str(sms_corpus)
```

```
## List of 5559
## $ 1 :List of 2
## ..$ content: chr "Hope you are having a good week. Just checking in"
## ..$ meta :List of 7
## ...$ author : chr(0)
## ...$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## ...$ description : chr(0)
## ...$ heading : chr(0)
## ...$ id : chr "1"
## ...$ language : chr "en"
## ...$ origin : chr(0)
## ...- attr(*, "class")= chr "TextDocumentMeta"
## ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 2 :List of 2
## ..$ content: chr "K..give back my thanks."
```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "2"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 3 :List of 2
## ..$ content: chr "Am also doing in cbe only. But have to pay."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "3"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 4 :List of 2
## ..$ content: chr "complimentary 4 STAR Ibiza Holiday or £10,000 cash needs your URGENT collection."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "4"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 5 :List of 2
## ..$ content: chr "okmail: Dear Dave this is your final notice to collect your 4* Tenerife Holiday o
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "5"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 6 :List of 2
## ..$ content: chr "Aiya we discuss later lar... Pick u up at 4 is it?"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "6"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 7 :List of 2
## ..$ content: chr "Are you this much buzy"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "7"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 8 :List of 2
## ..$ content: chr "Please ask mummy to call father"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "8"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 9 :List of 2
## ..$ content: chr "Marvel Mobile Play the official Ultimate Spider-man game (£4.50) on ur mobile ri
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "9"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 10 :List of 2
## ..$ content: chr "fyi I'm at usf now, swing by the room whenever"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "10"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 11 :List of 2
## ..$ content: chr "Sure thing big man. i have hockey elections at 6, shouldn't go on longer than an

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "11"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 12 :List of 2
## ..$ content: chr "I anything lor..."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "12"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 13 :List of 2
## ..$ content: chr "By march ending, i should be ready. But will call you for sure. The problem is t
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "13"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 14 :List of 2
## ..$ content: chr "Hmm well, night night "
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "14"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 15 :List of 2
## ..$ content: chr "K I'll be sure to get up before noon and see what's what"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "15"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 16 :List of 2
## ..$ content: chr "Ha ha cool cool chikku chikku:-):-DB-)"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "16"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 17 :List of 2
## ..$ content: chr "Darren was saying dat if u meeting da ge den we dun meet 4 dinner. Cos later u l
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "17"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 18 :List of 2
## ..$ content: chr "He dint tell anything. He is angry on me that why you told to abi."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "18"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 19 :List of 2
## ..$ content: chr "Up to u... u wan come then come lor... But i din c any stripes skirt..."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "19"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 20 :List of 2
## ..$ content: chr "U can WIN £100 of Music Gift Vouchers every week starting NOW Txt the word DRAW

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "20"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 21 :List of 2
## ..$ content: chr "2mro i am not coming to gym machan. Goodnight."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "21"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 22 :List of 2
## ..$ content: chr "ARR birthday today:) i wish him to get more oscar."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "22"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 23 :List of 2
## ..$ content: chr "Reading gud habit.. Nan bari hudgi yorge pataistha ertini kano:-)"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "23"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 24 :List of 2
## ..$ content: chr "I sent my scores to sophas and i had to do secondary application for a few school"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "24"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 25 :List of 2
## ..$ content: chr "Could you not read me, my Love ? I answered you"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "25"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 26 :List of 2
## ..$ content: chr "So what did the bank say about the money?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "26"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 27 :List of 2
## ..$ content: chr "Well if I'm that desperate I'll just call armand again"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "27"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 28 :List of 2
## ..$ content: chr "Fuuuuuck I need to stop sleepin, sup"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "28"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 29 :List of 2
## ..$ content: chr "So how's the weather over there?"

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "29"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 30 :List of 2
## ..$ content: chr "Ok thanx..."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "30"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 31 :List of 2
## ..$ content: chr "Ok.ok ok..then..whats ur todays plan"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "31"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 32 :List of 2
## ..$ content: chr "1Apple/Day=No Doctor. 1Tulsi Leaf/Day=No Cancer. 1Lemon/Day=No Fat. 1Cup Milk/day"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "32"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 33 :List of 2
## ..$ content: chr "Sorry, I'll call later"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "33"

```



```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 34 :List of 2
## ..$ content: chr "Will do. Was exhausted on train this morning. Too much wine and pie. You sleepw
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "34"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 35 :List of 2
## ..$ content: chr "U have won a nokia 6230 plus a free digital camera. This is what u get when u wi
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "35"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 36 :List of 2
## ..$ content: chr "Ron say fri leh. N he said ding tai feng cant make reservations. But he said wai
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "36"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 37 :List of 2
## ..$ content: chr "Call me when you/carlos is/are here, my phone's vibrate is acting up and I might
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "37"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 38 :List of 2
## ..$ content: chr "Oh k :)why you got job then whats up?"

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "38"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 39 :List of 2
## ..$ content: chr "SPJanuary Male Sale! Hot Gay chat now cheaper, call 08709222922. National rate f
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "39"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 40 :List of 2
## ..$ content: chr "Yeah you should. I think you can use your gt atm now to register. Not sure but i
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "40"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 41 :List of 2
## ..$ content: chr "Nationwide auto centre (or something like that) on Newport road. I liked them th
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "41"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 42 :List of 2
## ..$ content: chr "He is there. You call and meet him"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "42"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 43 :List of 2
## ..$ content: chr "Yeah sure I'll leave in a min"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "43"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 44 :List of 2
## ..$ content: chr "URGENT! Your Mobile number has been awarded with a £2000 prize GUARANTEED. Call 0
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "44"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 45 :List of 2
## ..$ content: chr "Mah b, I'll pick it up tomorrow"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "45"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 46 :List of 2
## ..$ content: chr "Then she dun believe wat?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "46"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 47 :List of 2
## ..$ content: chr "I've sent u my part.."

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "47"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 48 :List of 2
## ..$ content: chr "Hey so whats the plan this sat? "
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "48"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 49 :List of 2
## ..$ content: chr "Can you just come in for a sec? There's somebody here I want you to see"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "49"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 50 :List of 2
## ..$ content: chr "Yup. Izzit still raining heavily cos i'm in e mrt i can't c outside."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "50"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 51 :List of 2
## ..$ content: chr "I think we're going to finn's now, come"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "51"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 52 :List of 2
## ..$ content: chr "Our Prasanth ettans mother passed away last night. Just pray for her and family."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "52"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 53 :List of 2
## ..$ content: chr "Enjoy the showers of possessiveness poured on u by ur loved ones, bcoz in this w
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "53"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 54 :List of 2
## ..$ content: chr "He's really into skateboarding now despite the fact that he gets thrown off of i
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "54"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 55 :List of 2
## ..$ content: chr "You please give us connection today itself before DECIMAL or refund the bill"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "55"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 56 :List of 2
## ..$ content: chr "Congratulations ur awarded 500 of CD vouchers or 125gift guaranteed & Free entry

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "56"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 57 :List of 2
## ..$ content: chr "Mum not going robinson already."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "57"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 58 :List of 2
## ..$ content: chr "Got meh... When?"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "58"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 59 :List of 2
## ..$ content: chr "We are both fine. Thanks"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "59"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 60 :List of 2
## ..$ content: chr "awesome, how do I deal with the gate? Charles told me last night but, uh, yeah"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "60"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 61 :List of 2
## ..$ content: chr "I thought i'd get him a watch, just cos thats the kind of thing u get4an18th. And
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "61"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 62 :List of 2
## ..$ content: chr "Have a safe trip to Nigeria. Wish you happiness and very soon company to share m
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "62"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 63 :List of 2
## ..$ content: chr "Hey! do u fancy meetin me at 4 at cha - hav a lil beverage on me. if not txt or i
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "63"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 64 :List of 2
## ..$ content: chr "no, i *didn't* mean to post it. I wrote it, and like so many other times i've ri
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "64"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 65 :List of 2
## ..$ content: chr "If you want to mapquest it or something look up usf dogwood drive, that's the ti

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "65"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 66 :List of 2
## ..$ content: chr "Ha ha nan yalrigu heltini..Iyo kothi chikku, u shared many things wit me..so far
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "66"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 67 :List of 2
## ..$ content: chr "Becoz its # jan whn al the post office is in holiday so she cn go fr the post o
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "67"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 68 :List of 2
## ..$ content: chr "Good. do you think you could send me some pix? I would love to see your top and l
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "68"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 69 :List of 2
## ..$ content: chr "Beautiful Truth against Gravity.. Read carefully: Our heart feels light when some
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "69"

```



```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 70 :List of 2
## ..$ content: chr "Buzz! Hey, my Love ! I think of you and hope your day goes well. Did you sleep in"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "70"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 71 :List of 2
## ..$ content: chr "Lara said she can loan me # ."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "71"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 72 :List of 2
## ..$ content: chr "Dunno leh cant remember mayb lor. So wat time r we meeting tmr?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "72"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 73 :List of 2
## ..$ content: chr "So wats ur opinion abt him and how abt is character?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "73"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 74 :List of 2
## ..$ content: chr "Oh oh... Wasted... Den muz chiong on sat n sun liao..."

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "74"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 75 :List of 2
## ..$ content: chr "Are you at work right now ?"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "75"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 76 :List of 2
## ..$ content: chr "Bring home some Wendy =D"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "76"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 77 :List of 2
## ..$ content: chr "Ok going to sleep. Hope i can meet her."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "77"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 78 :List of 2
## ..$ content: chr "Allo! We have braved the buses and taken on the trains and triumphed. I mean we'"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "78"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 79 :List of 2
## ..$ content: chr "Hmmm:)how many players selected?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "79"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 80 :List of 2
## ..$ content: chr "You flippin out yet?"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "80"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 81 :List of 2
## ..$ content: chr "Your board is working fine. The issue of overheating is also reslove. But still s
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "81"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 82 :List of 2
## ..$ content: chr "URGENT! Your Mobile number has been awarded with a £2000 prize GUARANTEED. Call (
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "82"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 83 :List of 2
## ..$ content: chr "1000's of girls many local 2 u who r virgins 2 this & r ready 2 4fil ur every se

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "83"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 84 :List of 2
## ..$ content: chr "Nope. Since ayo travelled, he has forgotten his guy"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "84"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 85 :List of 2
## ..$ content: chr "Hey tmr maybe can meet you at yck"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "85"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 86 :List of 2
## ..$ content: chr "Ok i'm coming home now."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "86"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 87 :List of 2
## ..$ content: chr "u called dad oredi..."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "87"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 88 :List of 2
## ..$ content: chr "Alright, I'll head out in a few minutes, text me where to meet you"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "88"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 89 :List of 2
## ..$ content: chr "*deep sigh* ... I miss you :-( ... I am really surprised you haven't gone to the
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "89"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 90 :List of 2
## ..$ content: chr "Will be office around 4 pm. Now i am going hospital."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "90"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 91 :List of 2
## ..$ content: chr "ASKED 3MOBILE IF 0870 CHATLINES INCLU IN FREE MINS. INDIA CUST SERVs SED YES. L8"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "91"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 92 :List of 2
## ..$ content: chr "Yup bathe liao..."

```

```

## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "92"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 93 :List of 2
## ..$ content: chr "I am getting threats from your sales executive Shifad as i raised complaint again"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "93"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 94 :List of 2
## ..$ content: chr "S.this will increase the chance of winning."
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "94"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 95 :List of 2
## ..$ content: chr "How are you doing. How's the queen. Are you going for the royal wedding"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "95"
## .. ..$ language : chr "en"
## .. ..$ origin : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 96 :List of 2
## ..$ content: chr "Dear reached railway. What happen to you"
## ..$ meta :List of 7
## .. ..$ author : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description : chr(0)
## .. ..$ heading : chr(0)
## .. ..$ id : chr "96"

```

```

## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 97 :List of 2
## ..$ content: chr "Probably gonna swing by in a wee bit"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "97"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 98 :List of 2
## ..$ content: chr "K k pa Had your lunch aha."
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "98"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## $ 99 :List of 2
## ..$ content: chr "Whom you waited for yesterday"
## ..$ meta :List of 7
## .. ..$ author        : chr(0)
## .. ..$ timestamp: POSIXlt[1:1], format: "2020-04-23 05:56:53"
## .. ..$ description   : chr(0)
## .. ..$ heading       : chr(0)
## .. ..$ id            : chr "99"
## .. ..$ language      : chr "en"
## .. ..$ origin        : chr(0)
## .. ..- attr(*, "class")= chr "TextDocumentMeta"
## ..- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
## [list output truncated]
## - attr(*, "class")= chr [1:2] "VCorpus" "Corpus"

```

```

#print(sms_corpus)
inspect(sms_corpus[1:2])

```

```

## <<VCorpus>>
## Metadata: corpus specific: 0, document level (indexed): 0
## Content: documents: 2
##
## [[1]]
## <<PlainTextDocument>>
## Metadata: 7
## Content: chars: 49

```

```
##
## [[2]]
## <<PlainTextDocument>>
## Metadata: 7
## Content: chars: 23
```

```
as.character(sms_corpus[[1]])
```

```
## [1] "Hope you are having a good week. Just checking in"
```

```
lapply(sms_corpus[1:3], as.character)
```

```
## $`1`
## [1] "Hope you are having a good week. Just checking in"
##
## $`2`
## [1] "K..give back my thanks."
##
## $`3`
## [1] "Am also doing in cbe only. But have to pay."
```

```
# Lower the characters using transformation
sms_corpus_clean <- tm_map(sms_corpus,content_transformer(tolower))
as.character(sms_corpus[[1]])
```

```
## [1] "Hope you are having a good week. Just checking in"
```

```
as.character(sms_corpus_clean[[1]])
```

```
## [1] "hope you are having a good week. just checking in"
```

```
# Remove numbers
sms_corpus_clean <- tm_map(sms_corpus_clean,removeNumbers)
#> getTransformations()
# "removeNumbers" "removePunctuation" "removeWords" "stemDocument" "stripWhitespaces"
# Remove stopwords
sms_corpus_clean <- tm_map(sms_corpus_clean,removeWords,stopwords())
# Remove punctuations
sms_corpus_clean <- tm_map(sms_corpus_clean,removePunctuation)
# Stemming- reduce words to their root form
library(SnowballC)
wordStem(c("learn", "learned", "learning", "learns"))
```

```
## [1] "learn" "learn" "learn" "learn"
```

```
sms_corpus_clean <- tm_map(sms_corpus_clean,stemDocument)
# Remove white spaces
sms_corpus_clean <- tm_map(sms_corpus_clean,stripWhitespace)
as.character(sms_corpus[1:3])
```



```
## [1] "list(list(content = \"Hope you are having a good week. Just checking in\", meta = list(author =
## [2] \"list()\"
## [3] \"list()\"
```

```
as.character(sms_corpus_clean[1:3])
```

```
## [1] "list(list(content = \"hope good week just check\", meta = list(author = character(0), datetimes:
## [2] \"list()\"
## [3] \"list()\"
```

```
# Tokenization - split sentences to tokens
# Create Document Term Matrix - rows are documents and columns are terms
# We can also use Term Document Matrix which is a transpose of DTM
# Each cell of matrix is a number that represents the count of the word
# in the column in the document represented in the row
# Create DTM sparse matrix
sms_dtm <- DocumentTermMatrix(sms_corpus_clean)
sms_dtm2 <- DocumentTermMatrix(sms_corpus_clean, control = list(tolower = TRUE,
                                                                removeNumbers = TRUE,
                                                                stopwords = TRUE,
                                                                removePunctuation = TRUE,
                                                                stemming = TRUE
))
#sms_dtm
#sms_dtm2

# Create training and test datasets
sms_dtm_train <- sms_dtm[1:4169,]
sms_dtm_test <- sms_dtm[4170:5559,]
sms_train_labels <- sms_raw[1:4169,]$type
sms_test_labels <- sms_raw[4170:5559,]$type
# Compare the proportion of spam in training and test datasets
prop.table(table(sms_train_labels))
```

```
## sms_train_labels
##      ham      spam
## 0.8647158 0.1352842
```

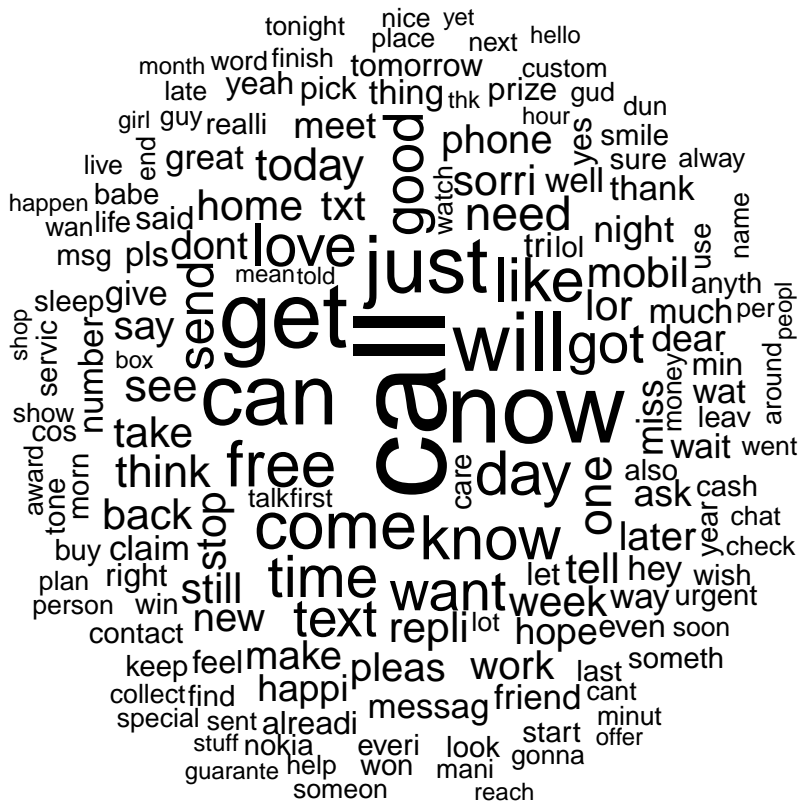
```
prop.table(table(sms_test_labels))
```

```
## sms_test_labels
##      ham      spam
## 0.8683453 0.1316547
```

```
# Visualize Text Data through Word Cloud
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

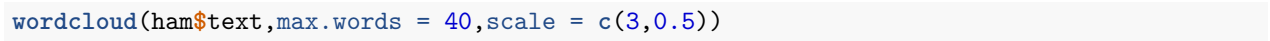
```
wordcloud(sms_corpus_clean,min.freq = 50,random.order = FALSE)
```



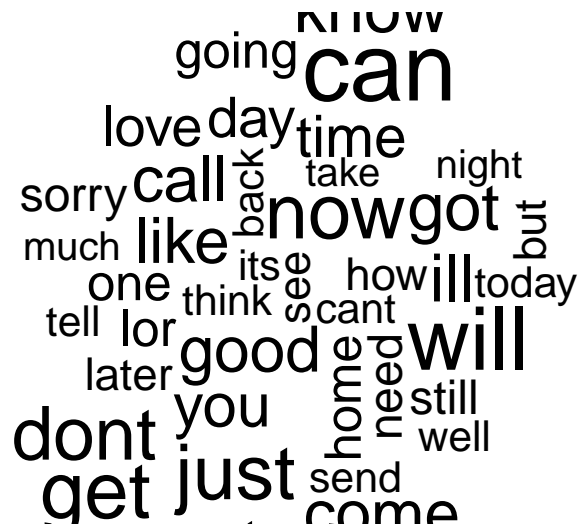
```
# Visualize spam and non-spam word clouds
spam <- subset(sms_raw,type == "spam")
ham <- subset(sms_raw,type == "ham")
wordcloud(spam$text,max.words = 40,scale = c(3,0.5))
```

```
## Warning in tm_map.SimpleCorpus(corpus, tm::removePunctuation): transformation
## drops documents
```

```
## Warning in tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x,
## tm::stopwords())): transformation drops documents
```



```
## Warning in tm_map.SimpleCorpus(corpus, tm::removePunctuation): transformation
## drops documents
```



```
# Create indicator features for Naive Bayes
str(sms_dtm)
```

```
## List of 6
## $ i      : int [1:42112] 1 1 1 1 1 2 2 2 3 3 ...
## $ j      : int [1:42112] 958 2270 2569 2924 6193 428 2976 5613 193 907 ...
## $ v      : num [1:42112] 1 1 1 1 1 1 1 1 1 1 ...
## $ nrow   : int 5559
## $ ncol   : int 6542
## $ dimnames:List of 2
## ..$ Docs : chr [1:5559] "1" "2" "3" "4" ...
## ..$ Terms: chr [1:6542] "'morrow" "'rent" "'llspeak" "'re" ...
## - attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
## - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

```
# Total 6542 features for words as number of columns
# Each word is a feature because it can help in detecting if an SMS is a spam or ham
# However, we eliminate words that appear in less than 5 messages or less than 0.1% of records
# in training data
# We use findFreqTerms function to find frequent words as a character vector
findFreqTerms(sms_dtm_train,5)
```

```
## [1] "&wk" "abiola" "abl"
## [4] "abt" "accept" "access"
## [7] "account" "across" "act"
```

##	[10]	"activ"	"actual"	"add"
##	[13]	"address"	"admir"	"adult"
##	[16]	"advanc"	"aft"	"afternoon"
##	[19]	"age"	"ago"	"aha"
##	[22]	"ahead"	"aight"	"aint"
##	[25]	"air"	"aiyo"	"alex"
##	[28]	"almost"	"alon"	"alreadi"
##	[31]	"alright"	"also"	"alway"
##	[34]	"angri"	"announc"	"anoth"
##	[37]	"answer"	"anymor"	"anyon"
##	[40]	"anyth"	"anytim"	"anyway"
##	[43]	"apart"	"app"	"appli"
##	[46]	"appreci"	"arcad"	"ard"
##	[49]	"area"	"argu"	"argument"
##	[52]	"armand"	"around"	"arrang"
##	[55]	"arriv"	"asap"	"ask"
##	[58]	"askd"	"attempt"	"auction"
##	[61]	"avail"	"ave"	"avoid"
##	[64]	"await"	"awak"	"award"
##	[67]	"away"	"awesom"	"babe"
##	[70]	"babi"	"back"	"bad"
##	[73]	"bag"	"bank"	"bare"
##	[76]	"basic"	"bath"	"batteri"
##	[79]	"bcoz"	"bday"	"beauti"
##	[82]	"becom"	"bed"	"bedroom"
##	[85]	"beer"	"begin"	"believ"
##	[88]	"best"	"better"	"bid"
##	[91]	"big"	"bill"	"bird"
##	[94]	"birthday"	"bit"	"black"
##	[97]	"blank"	"bless"	"blue"
##	[100]	"bluetooth"	"bold"	"bonus"
##	[103]	"boo"	"book"	"boost"
##	[106]	"bore"	"boss"	"bother"
##	[109]	"bout"	"box"	"boy"
##	[112]	"boytoy"	"break"	"breath"
##	[115]	"bring"	"brother"	"bslvyl"
##	[118]	"btnationalr"	"buck"	"bus"
##	[121]	"busi"	"buy"	"cabin"
##	[124]	"call"	"caller"	"callertun"
##	[127]	"camcord"	"came"	"camera"
##	[130]	"campus"	"can"	"cancel"
##	[133]	"cancer"	"cant"	"car"
##	[136]	"card"	"care"	"carlo"
##	[139]	"case"	"cash"	"cashbal"
##	[142]	"catch"	"caus"	"celebr"
##	[145]	"cell"	"centr"	"chanc"
##	[148]	"chang"	"charg"	"chat"
##	[151]	"cheap"	"cheaper"	"check"
##	[154]	"cheer"	"chennai"	"chikku"
##	[157]	"childish"	"children"	"choic"
##	[160]	"choos"	"christma"	"claim"
##	[163]	"class"	"clean"	"clear"
##	[166]	"close"	"club"	"code"
##	[169]	"coffe"	"cold"	"colleagu"

## [172]	"collect"	"colleg"	"colour"
## [175]	"come"	"comin"	"comp"
## [178]	"compani"	"competit"	"complet"
## [181]	"complimentari"	"comput"	"condit"
## [184]	"confirm"	"congrat"	"congratul"
## [187]	"connect"	"contact"	"content"
## [190]	"contract"	"cook"	"cool"
## [193]	"copi"	"correct"	"cos"
## [196]	"cost"	"cost&pm"	"costa"
## [199]	"coupl"	"cours"	"cover"
## [202]	"coz"	"crave"	"crazi"
## [205]	"creat"	"credit"	"cri"
## [208]	"cross"	"cuddl"	"cum"
## [211]	"cup"	"current"	"custcar"
## [214]	"custom"	"cut"	"cute"
## [217]	"cuz"	"dad"	"daddi"
## [220]	"darl"	"darlin"	"darren"
## [223]	"dat"	"date"	"day"
## [226]	"dead"	"deal"	"dear"
## [229]	"decid"	"decim"	"decis"
## [232]	"deep"	"definit"	"del"
## [235]	"deliv"	"deliveri"	"den"
## [238]	"depend"	"detail"	"didnt"
## [241]	"die"	"diet"	"differ"
## [244]	"difficult"	"digit"	"din"
## [247]	"dinner"	"direct"	"dis"
## [250]	"discount"	"discuss"	"disturb"
## [253]	"dnt"	"doc"	"doctor"
## [256]	"doesnt"	"dog"	"doin"
## [259]	"don"	"done"	"dont"
## [262]	"door"	"doubl"	"download"
## [265]	"draw"	"dream"	"drink"
## [268]	"drive"	"drop"	"drug"
## [271]	"dude"	"due"	"dun"
## [274]	"dunno"	"dvd"	"earli"
## [277]	"earlier"	"earth"	"easi"
## [280]	"eat"	"eatin"	"egg"
## [283]	"either"	"els"	"email"
## [286]	"embarass"	"end"	"energi"
## [289]	"england"	"enjoy"	"enough"
## [292]	"enter"	"entitl"	"entri"
## [295]	"envelop"	"etc"	"euro"
## [298]	"eve"	"even"	"ever"
## [301]	"everi"	"everybodi"	"everyon"
## [304]	"everyth"	"exact"	"exam"
## [307]	"excel"	"excit"	"excus"
## [310]	"expect"	"experi"	"expir"
## [313]	"extra"	"eye"	"face"
## [316]	"facebook"	"fact"	"fall"
## [319]	"famili"	"fanci"	"fantasi"
## [322]	"fantast"	"far"	"fast"
## [325]	"fat"	"father"	"fault"
## [328]	"feb"	"feel"	"felt"
## [331]	"fetch"	"fight"	"figur"

## [334]	"file"	"fill"	"film"
## [337]	"final"	"find"	"fine"
## [340]	"finger"	"finish"	"first"
## [343]	"fix"	"flag"	"flat"
## [346]	"flight"	"flower"	"follow"
## [349]	"fone"	"food"	"forev"
## [352]	"forget"	"forgot"	"forward"
## [355]	"found"	"freak"	"free"
## [358]	"freemsg"	"freephon"	"fren"
## [361]	"fri"	"friday"	"friend"
## [364]	"friendship"	"frm"	"frnd"
## [367]	"frnds"	"full"	"fullonsmscom"
## [370]	"fun"	"funni"	"fudur"
## [373]	"gal"	"game"	"gap"
## [376]	"gas"	"gave"	"gay"
## [379]	"gentl"	"get"	"gettin"
## [382]	"gift"	"girl"	"girlfrnd"
## [385]	"give"	"glad"	"god"
## [388]	"goe"	"goin"	"gone"
## [391]	"gonna"	"good"	"goodmorn"
## [394]	"goodnight"	"got"	"goto"
## [397]	"gotta"	"great"	"grin"
## [400]	"guarante"	"gud"	"guess"
## [403]	"guy"	"gym"	"haf"
## [406]	"haha"	"hai"	"hair"
## [409]	"half"	"hand"	"handset"
## [412]	"hang"	"happen"	"happi"
## [415]	"hard"	"hate"	"hav"
## [418]	"havent"	"head"	"hear"
## [421]	"heard"	"heart"	"heavi"
## [424]	"hee"	"hell"	"hello"
## [427]	"help"	"hey"	"hgsuiteland"
## [430]	"hit"	"hiya"	"hmm"
## [433]	"hmmm"	"hmv"	"hol"
## [436]	"hold"	"holder"	"holiday"
## [439]	"home"	"hook"	"hop"
## [442]	"hope"	"horni"	"hospit"
## [445]	"hot"	"hotel"	"hour"
## [448]	"hous"	"how"	"howev"
## [451]	"howz"	"hrs"	"httpwwwurawinnercom"
## [454]	"hug"	"huh"	"hungri"
## [457]	"hurri"	"hurt"	"ice"
## [460]	"idea"	"identifi"	"ignor"
## [463]	"ill"	"immedi"	"import"
## [466]	"inc"	"includ"	"india"
## [469]	"info"	"inform"	"insid"
## [472]	"instead"	"interest"	"invit"
## [475]	"ipod"	"irrit"	"ish"
## [478]	"island"	"issu"	"ive"
## [481]	"izzit"	"januari"	"jay"
## [484]	"job"	"john"	"join"
## [487]	"joke"	"joy"	"jst"
## [490]	"jus"	"just"	"juz"
## [493]	"kate"	"keep"	"kept"

## [496]	"kick"	"kid"	"kill"
## [499]	"kind"	"kinda"	"king"
## [502]	"kiss"	"knew"	"know"
## [505]	"knw"	"ladi"	"land"
## [508]	"landlin"	"laptop"	"lar"
## [511]	"last"	"late"	"later"
## [514]	"latest"	"laugh"	"lazi"
## [517]	"ldn"	"lead"	"learn"
## [520]	"least"	"leav"	"lect"
## [523]	"left"	"leh"	"lei"
## [526]	"less"	"lesson"	"let"
## [529]	"letter"	"liao"	"librari"
## [532]	"lie"	"life"	"lift"
## [535]	"light"	"like"	"line"
## [538]	"link"	"list"	"listen"
## [541]	"littl"	"live"	"lmao"
## [544]	"load"	"loan"	"local"
## [547]	"locat"	"log"	"lol"
## [550]	"london"	"long"	"longer"
## [553]	"look"	"lookin"	"lor"
## [556]	"lose"	"lost"	"lot"
## [559]	"lovabl"	"love"	"lover"
## [562]	"loyalti"	"ltd"	"luck"
## [565]	"lucki"	"lunch"	"luv"
## [568]	"mad"	"made"	"mah"
## [571]	"mail"	"make"	"malaria"
## [574]	"man"	"mani"	"march"
## [577]	"mark"	"marri"	"match"
## [580]	"mate"	"matter"	"maxim"
## [583]	"maxmin"	"may"	"mayb"
## [586]	"meal"	"mean"	"meant"
## [589]	"med"	"medic"	"meet"
## [592]	"meetin"	"meh"	"member"
## [595]	"men"	"merri"	"messag"
## [598]	"met"	"mid"	"midnight"
## [601]	"might"	"min"	"mind"
## [604]	"mine"	"minut"	"miracl"
## [607]	"miss"	"mistak"	"moan"
## [610]	"mob"	"mobil"	"mobileupd"
## [613]	"mode"	"mom"	"moment"
## [616]	"mon"	"monday"	"money"
## [619]	"month"	"morn"	"mother"
## [622]	"motorola"	"move"	"movi"
## [625]	"mrng"	"mrt"	"mrw"
## [628]	"msg"	"msgs"	"mths"
## [631]	"much"	"mum"	"murder"
## [634]	"music"	"must"	"muz"
## [637]	"nah"	"nake"	"name"
## [640]	"nation"	"natur"	"naughti"
## [643]	"near"	"need"	"net"
## [646]	"network"	"neva"	"never"
## [649]	"new"	"news"	"next"
## [652]	"nice"	"nigeria"	"night"
## [655]	"nite"	"nobodi"	"noe"

## [658]	"nokia"	"noon"	"nope"
## [661]	"normal"	"normpton"	"noth"
## [664]	"notic"	"now"	"num"
## [667]	"number"	"nyt"	"obvious"
## [670]	"offer"	"offic"	"offici"
## [673]	"okay"	"oki"	"old"
## [676]	"omg"	"one"	"onlin"
## [679]	"onto"	"oop"	"open"
## [682]	"oper"	"opinion"	"opt"
## [685]	"optout"	"orang"	"orchard"
## [688]	"order"	"oredi"	"oso"
## [691]	"other"	"otherwis"	"outsid"
## [694]	"pack"	"page"	"paid"
## [697]	"pain"	"paper"	"parent"
## [700]	"park"	"part"	"parti"
## [703]	"partner"	"pass"	"passion"
## [706]	"password"	"past"	"pay"
## [709]	"peopl"	"per"	"person"
## [712]	"pete"	"phone"	"photo"
## [715]	"pic"	"pick"	"pictur"
## [718]	"pin"	"piss"	"pix"
## [721]	"pizza"	"place"	"plan"
## [724]	"play"	"player"	"pleas"
## [727]	"pleasur"	"plenti"	"pls"
## [730]	"plus"	"plz"	"pmin"
## [733]	"pmsg"	"pobox"	"point"
## [736]	"poli"	"polic"	"poor"
## [739]	"pop"	"possess"	"possibl"
## [742]	"post"	"pound"	"power"
## [745]	"ppm"	"pray"	"present"
## [748]	"press"	"pretti"	"previous"
## [751]	"price"	"princess"	"privat"
## [754]	"prize"	"prob"	"probabl"
## [757]	"problem"	"project"	"promis"
## [760]	"pub"	"put"	"qualiti"
## [763]	"question"	"quick"	"quit"
## [766]	"quiz"	"quot"	"rain"
## [769]	"random"	"rang"	"rate"
## [772]	"rather"	"rcvd"	"reach"
## [775]	"read"	"readi"	"real"
## [778]	"reali"	"realli"	"reason"
## [781]	"receipt"	"receiv"	"recent"
## [784]	"record"	"refer"	"regard"
## [787]	"regist"	"relat"	"relax"
## [790]	"remain"	"rememb"	"remind"
## [793]	"remov"	"rent"	"rental"
## [796]	"repli"	"repres"	"request"
## [799]	"respond"	"respons"	"rest"
## [802]	"result"	"return"	"reveal"
## [805]	"review"	"reward"	"right"
## [808]	"ring"	"rington"	"rite"
## [811]	"road"	"rock"	"role"
## [814]	"room"	"roommat"	"rose"
## [817]	"round"	"rowwjhl"	"rpli"

## [820]	"rreveal"	"run"	"rush"
## [823]	"sad"	"sae"	"safe"
## [826]	"said"	"sale"	"sat"
## [829]	"saturday"	"savamob"	"save"
## [832]	"saw"	"say"	"sch"
## [835]	"school"	"scream"	"sea"
## [838]	"search"	"sec"	"second"
## [841]	"secret"	"see"	"seem"
## [844]	"seen"	"select"	"self"
## [847]	"sell"	"semest"	"send"
## [850]	"sens"	"sent"	"serious"
## [853]	"servic"	"set"	"settl"
## [856]	"sex"	"sexi"	"shall"
## [859]	"share"	"shd"	"ship"
## [862]	"shirt"	"shop"	"short"
## [865]	"show"	"shower"	"sick"
## [868]	"side"	"sigh"	"sight"
## [871]	"sign"	"silent"	"simpl"
## [874]	"sinc"	"singl"	"sipix"
## [877]	"sir"	"sis"	"sister"
## [880]	"sit"	"situat"	"skxh"
## [883]	"skype"	"slave"	"sleep"
## [886]	"slept"	"slow"	"slowli"
## [889]	"small"	"smile"	"smoke"
## [892]	"sms"	"smth"	"snow"
## [895]	"sofa"	"sol"	"somebodi"
## [898]	"someon"	"someth"	"sometim"
## [901]	"somewher"	"song"	"soni"
## [904]	"sonyericsson"	"soon"	"sorri"
## [907]	"sort"	"sound"	"south"
## [910]	"space"	"speak"	"special"
## [913]	"specialcal"	"spend"	"spent"
## [916]	"spoke"	"spree"	"stand"
## [919]	"start"	"statement"	"station"
## [922]	"stay"	"std"	"step"
## [925]	"still"	"stockport"	"stone"
## [928]	"stop"	"store"	"stori"
## [931]	"street"	"student"	"studi"
## [934]	"stuff"	"stupid"	"style"
## [937]	"sub"	"subscrib"	"success"
## [940]	"suck"	"suit"	"summer"
## [943]	"sun"	"sunday"	"sunshin"
## [946]	"sup"	"support"	"suppos"
## [949]	"sure"	"surf"	"surpris"
## [952]	"sweet"	"swing"	"system"
## [955]	"take"	"talk"	"tampa"
## [958]	"tariff"	"tcs"	"tea"
## [961]	"teach"	"tear"	"teas"
## [964]	"tel"	"tell"	"ten"
## [967]	"tenerif"	"term"	"test"
## [970]	"text"	"thank"	"thanx"
## [973]	"that"	"thing"	"think"
## [976]	"thinkin"	"thk"	"tho"
## [979]	"though"	"thought"	"throw"

## [982]	"thru"	"tht"	"thur"
## [985]	"tick"	"ticket"	"til"
## [988]	"till"	"time"	"tire"
## [991]	"titl"	"tmr"	"toclaim"
## [994]	"today"	"togeth"	"told"
## [997]	"tomo"	"tomorrow"	"tone"
## [1000]	"tonight"	"tonit"	"took"
## [1003]	"top"	"torch"	"tot"
## [1006]	"total"	"touch"	"tough"
## [1009]	"tour"	"toward"	"town"
## [1012]	"track"	"train"	"transact"
## [1015]	"travel"	"treat"	"tri"
## [1018]	"trip"	"troubl"	"true"
## [1021]	"trust"	"truth"	"tscs"
## [1024]	"ttyl"	"tuesday"	"turn"
## [1027]	"twice"	"two"	"txt"
## [1030]	"txting"	"txts"	"type"
## [1033]	"ufind"	"ugh"	"ull"
## [1036]	"uncl"	"understand"	"unless"
## [1039]	"unlimit"	"unredeem"	"unsub"
## [1042]	"unsubscribe"	"updat"	"ure"
## [1045]	"urgent"	"urself"	"use"
## [1048]	"user"	"usf"	"usual"
## [1051]	"uve"	"valentin"	"valid"
## [1054]	"valu"	"via"	"video"
## [1057]	"vikki"	"visit"	"vodafon"
## [1060]	"voic"	"vomit"	"voucher"
## [1063]	"wait"	"wake"	"walk"
## [1066]	"wan"	"wana"	"wanna"
## [1069]	"want"	"wap"	"warm"
## [1072]	"wast"	"wat"	"watch"
## [1075]	"water"	"way"	"weak"
## [1078]	"wear"	"weather"	"wed"
## [1081]	"wednesday"	"weed"	"week"
## [1084]	"weekend"	"welcom"	"well"
## [1087]	"wen"	"went"	"what"
## [1090]	"whatev"	"whenev"	"whole"
## [1093]	"wid"	"wif"	"wife"
## [1096]	"wil"	"will"	"win"
## [1099]	"wine"	"winner"	"wish"
## [1102]	"wit"	"within"	"without"
## [1105]	"wiv"	"wkli"	"wks"
## [1108]	"wnt"	"woke"	"won"
## [1111]	"wonder"	"wont"	"word"
## [1114]	"work"	"workin"	"world"
## [1117]	"worri"	"wors"	"worth"
## [1120]	"wot"	"wow"	"write"
## [1123]	"wrong"	"wwq"	"wwwgetzedcouk"
## [1126]	"xmas"	"xxx"	"yahoo"
## [1129]	"yar"	"yeah"	"year"
## [1132]	"yep"	"yes"	"yesterday"
## [1135]	"yet"	"yoga"	"yup"

```
sms_freq_words <- findFreqTerms(sms_dtm_train,5)
str(sms_freq_words)
```

```
## chr [1:1137] "&wk" "abiola" "abl" "abt" "accept" "access" "account" ...
```

```
# We filter the words in document term matrix that appear 5 times or more
sms_freq_dtm_train <- sms_dtm_train[,sms_freq_words]
sms_freq_dtm_test <- sms_dtm_test[,sms_freq_words]
str(sms_freq_dtm_train)
```

```
## List of 6
## $ i      : int [1:24942] 1 1 1 1 1 2 2 3 3 4 ...
## $ j      : int [1:24942] 153 392 442 491 1083 71 971 32 708 140 ...
## $ v      : num [1:24942] 1 1 1 1 1 1 1 1 1 1 ...
## $ nrow    : int 4169
## $ ncol    : int 1137
## $ dimnames:List of 2
## ..$ Docs : chr [1:4169] "1" "2" "3" "4" ...
## ..$ Terms: chr [1:1137] "&wk" "abiola" "abl" "abt" ...
## - attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
## - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

```
str(sms_freq_dtm_test)
```

```
## List of 6
## $ i      : int [1:7981] 1 1 1 2 2 2 2 2 3 ...
## $ j      : int [1:7981] 175 247 439 131 175 723 735 814 1084 192 ...
## $ v      : num [1:7981] 1 1 1 1 1 1 1 1 1 1 ...
## $ nrow    : int 1390
## $ ncol    : int 1137
## $ dimnames:List of 2
## ..$ Docs : chr [1:1390] "4170" "4171" "4172" "4173" ...
## ..$ Terms: chr [1:1137] "&wk" "abiola" "abl" "abt" ...
## - attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
## - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

```
# We finally have 1137 terms which will act as features for Naive Bayes
# Naive Bayes is trained on categorical variables, so we will change the terms frequency
# to "Yes" or "No" i.e. whether a term appears or not
# This approach loses the important information about the frequency of words
convert_counts <- function(x) { x <- ifelse(x>0, "Yes", "No")}
convert_counts(c(5,3,7))
# We use apply function with margin 2 (column) to apply convert_counts
sms_train <- apply(sms_freq_dtm_train,MARGIN = 2,convert_counts)
sms_test <- apply(sms_freq_dtm_test,MARGIN = 2,convert_counts)
#sms_train
#sms_test
str(sms_train)
```

```
## chr [1:4169, 1:1137] "No" "No" "No" "No" "No" "No" "No" "No" "No" "No" ...
## - attr(*, "dimnames")=List of 2
## ..$ Docs : chr [1:4169] "1" "2" "3" "4" ...
## ..$ Terms: chr [1:1137] "&wk" "abiola" "abl" "abt" ...
```

```
#sms_train[which(sms_train=="Yes")]
#sms_test[which(sms_test=="Yes")]
table(sms_train)
```

```
## sms_train
##      No      Yes
## 4715211 24942
```

```
table(sms_test)
```

```
## sms_test
##      No      Yes
## 1572449 7981
```

```
# Use Naive Bayes implementation from e1071 package by Viena Institute of Technology
library(e1071)
```

```
# We can also use NaiveBayes from klaR package
```

```
#Train the classifier
```

```
sms_classifier <- naiveBayes(sms_train,sms_train_labels)
```

```
# Predict on test set
```

```
sms_test_pred <- predict(sms_classifier,sms_test)
```

```
# Evaluation
```

```
library(gmodels)
```

```
CrossTable(sms_test_pred,sms_test_labels,prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE, dnn = c('p
```

```
##
```

```
##
```

```
##      Cell Contents
```

```
## |-----|
## |                      N |
## |          N / Table Total |
## |-----|
```

```
##
```

```
##
```

```
## Total Observations in Table: 1390
```

```
##
```

```
##
```

```
##      | actual
## predicted |      ham |      spam | Row Total |
## -----|-----|-----|-----|
##      ham |      1201 |        30 |      1231 |
##      |      0.864 |      0.022 |      |
## -----|-----|-----|-----|
##      spam |         6 |       153 |       159 |
##      |      0.004 |      0.110 |      |
## -----|-----|-----|-----|
## Column Total |      1207 |       183 |      1390 |
## -----|-----|-----|-----|
```

```
##
```

```
##
```

```
table(sms_test_pred)
```

```
## sms_test_pred
##   ham spam
## 1231  159
```

```
table(sms_test_labels)
```

```
## sms_test_labels
##   ham spam
## 1207  183
```

```
# We note
# Our classifier predicted 1231 ham whereas only 1201 were actually ham, so 30 spam messages
# were incorrectly reported as ham
# Our classifier predicted 159 messages as spam wherein only 153 were actually spam and 6 ham
# messages were incorrectly reported as spam
# Total incorrectly classified messages: 6+30 = 36
# We can improve model performance by setting Laplace smoothing parameter to 1
# Train the classifier with Laplace 1
sms_classifier <- naiveBayes(sms_train,sms_train_labels,laplace = 1)
# Predict on test set
sms_test_pred <- predict(sms_classifier,sms_test)
# Evaluation
library(gmodels)
CrossTable(sms_test_pred,sms_test_labels,prop.chisq = FALSE, prop.c = FALSE, prop.r = FALSE, dnn = c('p'
```

```
##
##
##   Cell Contents
## |-----|
## |                      N |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table:  1390
##
##
##      | actual
## predicted |      ham |      spam | Row Total |
## -----|-----|-----|-----|
##      ham |      1202 |         28 |        1230 |
##      |      0.865 |      0.020 |           |
## -----|-----|-----|-----|
##      spam |         5 |        155 |         160 |
##      |      0.004 |      0.112 |           |
## -----|-----|-----|-----|
## Column Total |      1207 |         183 |        1390 |
## -----|-----|-----|-----|
##
##
```

```
table(sms_test_pred)
```

```
## sms_test_pred  
##   ham spam  
## 1230  160
```

```
table(sms_test_labels)
```

```
## sms_test_labels  
##   ham spam  
## 1207  183
```

```
# We notice that our model performance worsened. Therefore, Laplace smoothing did not help.
```