

Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language

Ahmed Imran KABIR¹, Koushik AHMED², Ridoan KARIM³

¹ Adjunct Lecturer, School of Business and Economics, United International University, Bangladesh.

² Undergraduate Student, School of Business and Economics, United International University, Bangladesh.

³ Lecturer, School of Business, Department of Business Law and Taxation, Monash University, Malaysia.

ahmedimran@bus.uui.ac.bd, koushikahmed.x1@gmail.com, ridoan.karim@monash.edu

The internet has opened a very wide range of ways for exchanging information or data. The development of internet influenced our daily lives to share our opinion on internet. We can share our opinions in social media like twitter, Facebook, LinkedIn or micro blogging site. We can give reviews about any product or we can share what things we are expecting. The sharing of information in internet makes internet a rich resource. Large organization or any type of business who want to do business in a customer centric way needs to know what peoples are thinking. To know that, we can use online resources but to analyze all the data in a short time is not easy if we try to figure out everyone is thought one by one. Sentiment analysis and word cloud in text mining is introduced to eradicate this problem. It helps to know what peoples are thinking and helps to develop the client experience and helps to take decision in a customer centric way. The project on word cloud and sentiment analysis of amazon earphones reviews is basically done to know the process which we can used in our practical life to know the people's attitude, opinions, reviews, sentiment towards something from unstructured big data from online resources.

Keywords: Word Cloud, Sentiment Analysis, Big Data Analysis, R Programming Language

DOI: 10.24818/issn14531305/24.4.2020.05

1 Introduction

Internet technology and its uses are growing rapidly. Because of internet, uses of social site such as Facebook, twitter, Instagram, Tumblr and other micro blogging site or business website is increasing. People share their thoughts, sentiments, attitudes and reviews on online. That is why online resources are getting rich day by day. As a business student it is more important to know what consumers are thinking. As information technology has developed, it is getting easy to understand our customer through online resources where they share their thoughts, attitudes, opinion, reviews and sentiments. So, understanding our customer or people through their text will be discussed here. This is the reason to work with word cloud and sentiment analysis which helps us to understand our customers or people thinking's, sentiments, attitudes, opinions etc.

To understand people's attitudes, sentiments, and opinions from text is very difficult because of the unstructured text and a dataset can have millions of information from people. So, to analyze big data and come up with their attitude is not easy. So here we will try to figure out people's attitude from their unstructured text.

The main objectives of this study are:

- Learning natural text processing
- Learning data cleaning
- Making structured data from unstructured data
- Making bar plot
- Making word cloud and word cloud 2
- Learning sentiment analysis
- Determining positive, negative or neutral attitude from sentiment analysis.

This paper describes natural text processing, data cleaning, word cloud making and

sentiment analysis of amazon earphones reviews. The aim is to find out opinions from amazon earphones reviews are positive, neutral or negative. From word cloud it describes those words what people use more in their reviews.

While preparing this report we have faced some limitations to complete this paper and they are given below: -

Unstructured data: Peoples writings on website or social site are mainly unstructured text, sometimes they use special character which cannot be easily found from a large data set.

2 Literature Review

In the 21st century Internet technology is growing bigger and it is producing many opportunities for people. We can see that in nowadays most people are providing their information, thoughts, reviews about something on social media like Facebook, twitter, Instagram etc. micro blogging site, business website or personal website. People are making valuable source of online perception. The bigger industries and firms take advantages of people's feedback to improve their products and services and it will help them to enhance their marketing strategies. One example can be given which Apple iPhone actually create a hype for upcoming iPhone to extract people's emotions and market the product before its releases [1]. There are many ways to analysis or extract people's emotions through online data and it will be discussed here that how we can get their perception about something.

The R Language

R is a programming language which is used for statistical analysis and graphics. It also gives debugging facilities. This manual detail and defines the R language. In the 1980s the R language was designed. Form then It was used in statistical society. The main designer of R is John M. Chambers. Using command line, it is possible to execute simple expressions. Users do not need to always create function because there are lots of add on available in the R language for statistical

analysis. There are many shortcuts and idioms which helps user to do critical analysis simply. There are many ways of doing a specific task but some of them depends on the language[2].

Corpus

For managing documents in text mining is known as corpus which actually represents the collection of documents[3]. For conducting a study of corpus-based language it is important to gain access to corpus. It consists of dataset of natural texts compiled from writings or transcription of recorded speech. The main purpose of corpus is to find the pattern of natural languages and its behavior. It gives a more objective view of language than that of introspection, intuition and anecdotes. A corpus-based analysis can inspect almost any language patterns--lexical, structural, lexicon-grammatical, discourse, phonological, morphological--often with very specific agendas[4].

Data Cleaning

Data cleaning or data preprocessing is an important role for any statistical analysis. It is more time consuming. There are many techniques in R statistical environment which helps to build data cleaning script[5]. There are some common steps for data cleaning are given below: -

- Making everything to lowercase
- Removing punctuation from the dataset
- Removing all numbers from the dataset
- Removing stop words from the dataset
- Removing specific word which you do not want in the dataset
- Removing URL or website link from the dataset
- Removing extra white space

Term Document Matrix

Term document matrix basically presents the words from dataset as text in the table of numbers. Term document matrix presents the words into rows and columns. The rows show the words which need to be analyzed and the columns show the appearance of words. Binary version is used here. A 1 indicates the

existence and a 0 represent absence of a word[6].

Sparse matrix

A sparse matrix gives more well-organized view of the data contained in the term document matrix. When there are large numbers of words or cases, it is important to use sparse matrix. The matrix shows us how much zeroes exists in this data proportionately. It can save computer memory as it does not store information about the presentation. The problem of this matrix is it does not show as nicely on the screen. So you need to convert into a normal matrix if you want to do further calculation or analysis[6].

Bar Plot

The *barplot()* function can create bar plots using R. A vector or matrix can be supplied into this function. The plot will create bar heights equal to the elements if we provide vector. As we know vector as well as matrix can be taken by *barplot()*. So, if we provide matrix as input, it will create stacked bar. A stacked bar will represent each column of the matrix.

Word Cloud

Word cloud is basically a visually representation of text. Word clouds are used in various ways. Generally, this can be made with pure text summarization. Word clouds are mainly used in text analytics. Word clouds are mainly made with a body of text. Word clouds help to know that how much similar an information is for specific research. They give summery of isolated words without knowing their linguistic meaning or relations. They are statistically used and provide no or limited interaction capabilities [7]. Word cloud 2, letter cloud are the other versions of word clouds with distinct feature. Word cloud 2 gives traditional word cloud with HTML5 and letter cloud prove a word cloud with selected letters or word (Figure 1).



Fig. 1. Word Cloud Example

Sentiment Analysis

The main purpose of information gathering is to know what other people think. To know what other people think there are many approaches and one of them is sentiment analysis or opinion mining. The opinion resources are growing day by day because of availability of internet. People tend to give their opinion in online such as reviews of any product or service, election opinion, writings in micro blog, gives feedback etc. As online resources are getting rich day by day, Information technology can be used to know the understandings of others people [8]. Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity [9]. Sentiment analysis is the field of analyzing people's attitudes, sentiments, evaluations, and emotions from text. Sentiment analysis is all about natural language processing and it is also used in web mining, text mining and data mining. Formerly it was studied in computer science but recently it is used in Management Information Systems and Business because It helps a lot in business to take decisions according to people's perception [10].

Legal and Ethical Aspects of Sentiment Analysis

Sentiment analysis deals with the feeling, ideas and actions of written texts. Using algorithms to discern the feelings behind words, sentiment analysis will decide whether a message indicates a positive, negative or neutral response. Although the technology is nothing new, studies on the interpretation of feelings have raised serious questions about the technology's legal and ethical stands [11].

Of course, this is becoming less of a problem to understand the legal issues because of the implementation of GDPR; however, corporations aiming for sentiment analyses should consider additional legal perspectives that indirectly regulate sentiment analysis.

Data from social networking are rich sources of knowledge to understand the consumer's values and customer diversity in the global marketplace [12]. Moreover, political parties vying for state elections may also find social media study very useful for understanding the people's viewpoint and designing new political strategies for electoral events. There are some major legal issues sentiment analysis as several data sets are protected and governed by the database law, copyright law and contract law [13]. The copyright law gives protection automatically over the work which one not copied from elsewhere by recording and creating in a way it establishes a new inventory work [13]. The copyright protection somehow provides the owner the right whether to approve or decline on third party's' restricts acts which includes such as adaptation of the work, copying, re-dissemination for publishing it on the Web, or any substantial part of the original copyright work, as well as translation of the work into other languages.

Besides, there are certain ethical considerations that could be identified in the ways in which companies access sentiment, use sentiment, and intervene in sentiment. There are obviously less legal and ethical issues if sentiment analysis is conducted based on publicly available data. However, it is well established that what is public and private is complex in social media environments. Several researchers use the notion of "being private in public" to highlight this complexity, comparing social media with corridors [14]. If two people bump into each other in the public space of a corridor, she argues, one might say something private to the other that s/he would not want to have publicized. Instead, the speaker would want "privacy in context," or what is known as "contextual integrity."

Understanding private or public domains of data are becoming more explicitly critical. Many companies using sentiment analysis, benefit both from the opacity of what is public in social media. Lack of appropriate legislation, and the absence of legal liability has become a great excuse for failing to exercise due diligence. Thus, a number of actors are invoked as having responsibilities in relation to sentiment analysis.

Because of the opportunities available to engage with and get heard by companies and brands, consumers now have more control over brands. People have more power to impact on how companies behave as the companies actively involve themselves in sentiment analysis. Corporations running sentiment analysis unethically contains the risk of negatively impacting individuals or groups of people. In most of the countries, the constitution ensures that neither the Government nor a large corporation can harm a single individual without 'the due process of law'.

Consumers should be properly aware as to what sentiment analysis is and how it can be used to better their shopping experience. For example, customer complaints management is a simple example how sentiment analysis on both sides of the purchasing and selling can help customers. Latest surveys show that 67 percent of customers have used customer services from a company's social media page [15]. In reality, 33% prefer this telephone communication channel [15].

This is where sentiment analysis can benefit both business and customer. Using the technology, tweets that express a particularly negative sentiment about the company can be instantly highlighted for the attention of management. That might be an incorrect delivery complaint, a problem with a product, or anything containing inflammatory language.

3 Research Methodology

In the 21st century, it is important to know people's perception. So, we will describe how we can get a closure look of people's perception according to their online reviews

of earphones. We will figure out their attitudes, emotions and evaluations based on their online resources.

Research Design

Research design is a framework which a researcher used to do analysis. The appropriate design can make a successful research. There are many types of research design but most of them follows some basic steps and they are:

1. Data collection
2. Data Measurement
3. Data analysis
4. My research design is given below step by step-
5. Data collection (Raw Data)
6. Data preprocessing
7. Data Cleaning
8. Term Document Matrix
9. Word cloud
10. Obtain score
11. Sentiment analysis
12. Visualization

Data Collection

According to, [16] Data collection is the process of collecting data from sources. There are mainly two types of data collection method and they are primary data and secondary data.

Primary data are data that are collected for specific research using survey, interview, field observation, and experiments. It is expensive and time consuming.

Secondary data are already collected or used for another task. It can be collected from newspapers, journals, annual reports etc.

We used here secondary data which is all about Amazon earphones reviews and the data is collected from the Kaggle website. The dataset link is:

<https://www.kaggle.com/shitalkat/amazonearphonesreviews>.

Sample

Sample is a subset of total data. As we used amazon earphones reviews data sets that means it should include all type of brands and all reviews. But it's expensive and time consuming. So that we choose these brands which includes total 14338 reviews, and this is our sample.

- boAt Rockerz 255
- Flybot Wave
- Flybot Boom
- PTron Intunes
- Flybot Beat
- Samsung EO-BG950CBEIN
- JBL T205BT
- Sennheiser CX 6.0BT
- Skullcandy S2PGHW-174
- JBL T110BT

Research process

Here we are going to describe the steps to create word cloud and sentiment analysis steps. In this work we used R programming language and R Studio IDE. The steps of analysis are given below.

Read File

To read csv file we use this code: `earphones <- read.csv(file.choose(), header = T)`. As our csv file is all about earphone reviews so we store this code: `read.csv(file.choose(), header = T)` in earphones. "header=T" indicates that the CSV file has header that means each column has title. When we ran this code a popup window is opened to read our CSV file and we choose our file which name was "AllProductReviews.csv" (Figure 2).

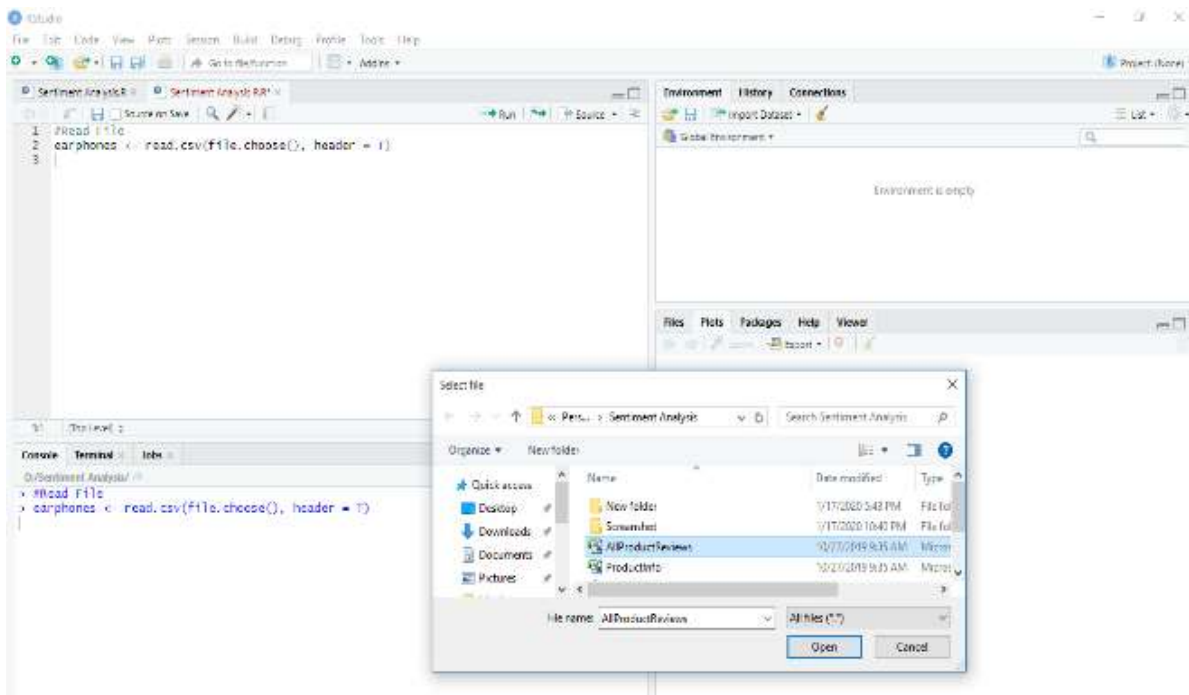


Fig. 2. Reading File in R Studio

- After reading our file we can see that in the right side the data shows that there are 14,337 observations and 4 variables. That means there are 14,337 rows of reviews and four columns. So, we want to look at the structure of our file.
- To know the structure of the file we ran this code - `str(earphones)`. It shows the name of four variables and they are `ReviewTitle`, `ReviewBody`, `ReviewStar` and `Product`. Here our only interest is on `ReviewBody` that contains reviews about earphones.

Build Corpus

- Corpus is basically a collection of documents. So, in this case each review is

treated as document. First we use the code – `library [17]` where `tm` refers to text mining.

- Before doing anything we use the code: `corpus <- iconv(earphones$ReviewBody, to = 'utf8')`, where `Reviewbody` indicates our interested area and `utf8` is the format of the text.

After that we ran this code: `corpus <- Corpus (VectorSource(corpus))` to create corpus and to store it in the corpus. We inspected this corpus with this code: `inspect (corpus [1:5])`. This code inspects first five reviews (Figure 3).

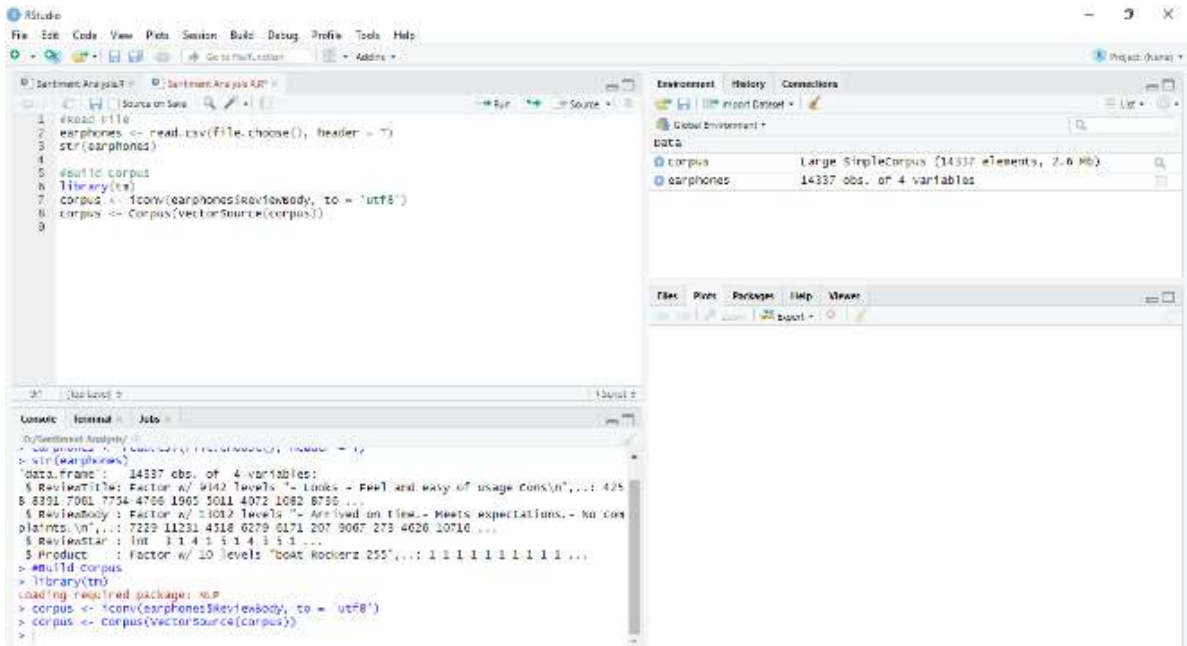


Fig. 3. Build Corpus in R Studio

Data Cleaning

- To make everything lowercase we ran this code: `corpus <- tm_map(corpus, tolower)`.
- To inspect lowercase again we inspected first five reviews with this code: `inspect(corpus[1:5])`. It shows that every text in that document turned into lowercase.
- Then we removed all punctuations like full stops, commas and store this in corpus with this code: `corpus <- tm_map(corpus, removePunctuation)`.
- Then again we inspected this corpus to ensure that every punctuation is removed with this code: `inspect(corpus[1:5])`.
- Then we remove all the numbers with this code: `corpus <- tm_map(corpus, removeNumbers)` and store it in the corpus.
- Then we checked it again with this code: `inspect(corpus[1:5])` and it shows that all numbers are gone from this document.
- Then we remove some English words that are so common and not adding much value. We can specify them as stop words so we remove them from this code: `cleanset <- tm_map(corpus, removeWords, stopwords('english'))` and store it in cleanset.

- To know what stop words looks like we can select `stopwords('english')` and run this. There are total 174 stop words.
- Again we inspect this `cleanset` with this code: `inspect(cleanset[1:5])` and it shows that all stop words in English are gone.
- We can see that there is “/n” in the end of every reviews. This is not adding any value so we remove this using function. The code is:
`removeLast <- function(x) gsub('\n', '', x)`
`cleanset <- tm_map(cleanset, content_transformer(removeLast))`
- Again we inspected first five reviews to ensure that “/n” is gone from every reviews with this code: `inspect(cleanset[1:5])` and it shows “/n” is gone from all documents.
- After that we remove website link from all reviews using function with this code:
`removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)`
`cleanset <- tm_map(cleanset, content_transformer(removeURL))`
- Again we inspected first five reviews with this: `inspect(cleanset[1:5])`
- Another thing coder may notice is as when something is removed so it leaves a white space. So as we wanted to get rid of that too, we use a code to remove extra white

space and the code is: `cleanset <- tm_map(cleanset, stripWhitespace)` and stored it in a `cleanset`.

- Then again we inspected this using the code: `inspect(cleanset[1:5])` and it shows all extra white spaces are gone.

Term Document Matrix

Text data like reviews are basically unstructured data. So to do any further analysis we need to convert this data into a structured data into rows and columns. That is achieved by term document matrix.

- To create term document matrix, we ran this code: `tdm <- TermDocumentMatrix(cleanset)`
- Then we ran `tdm` and it gives some information about the structured data.
- From that we can see that there are 14746 terms that means words and 14337 documents means the number of reviews. This is a huge matrix. Here sparsity is 100% so we can see 100% zeroes in this matrix.
- We want to turn into this as matrix so we have to run this following code: `tdm <- as.matrix(tdm)`
- We want to get an idea how it looks so we ran this code for first 10 words and 1 to 20 columns: `tdm[1:10, 1:20]`

Bar Plot

- We can find how often each word appears in this dataset by doing row sums of `tdm`.

- So, we ran this code: `w <- rowSums(tdm)` then select `w` and run and it gives the frequency of each words in this data set.

- As we have a large dataset with 14746 words and 14337 number of reviews so that we want to take a subset from this dataset which frequency is equal or more than 300 which will be stored in `w`. So we ran this code: `w <- subset(w, w>=300)`

- Then again, we select `w` and ran again to know how it looks and it only shows those words that have equal or more than 300 frequency.

- Then we create a bar plot with this subset. For creating bar plot we use this code-

```
barplot(w,
        las=2,
        col = rainbow(50))
```

- Here `w` represents the subset which have equal to or more than 300 frequency, `las=2` represents all the words are listed vertically on x axis and color used here is rainbow.

- We can see that there are two words earphones and earphone. As this data set is all about earphones so that we do not want these two words in our bar plot otherwise they will overshadow the dataset. So we go back to `cleanset` and ran this code: `cleanset <- tm_map(cleanset, removeWords, c('earphone', 'earphones', 'boat'))`

- Then again, we ran all these codes to get a bar plot and the bar plot is looks like below (Figure 4)

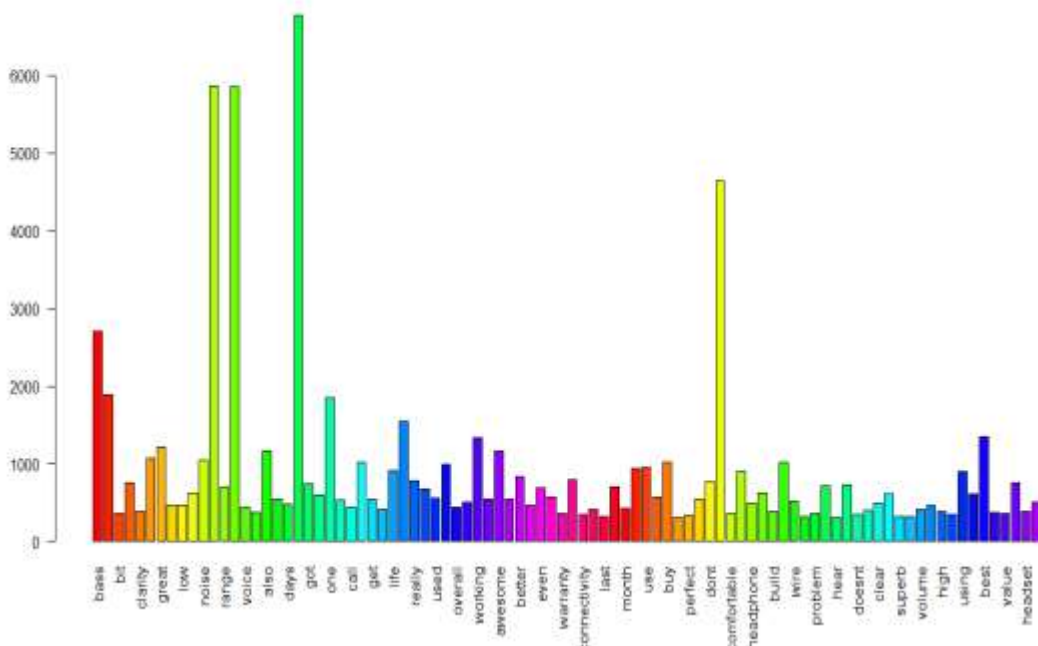


Fig. 4. Bar Plot Creation in R Studio

Word Cloud

Word cloud is basically an electronic image that shows words in a particular piece of electronic text or series of text. Here words are different sizes according to their frequency.

- For creating word cloud first, we need a library word cloud so that we use a library first.
- Then we create another `w` and sort this as per frequency of `rowSums` and sort this with decreasing order by using this code: `w <- sort(rowSums(tdm), decreasing = TRUE)`
- For repeatability we used this code- `set.seed(222)`
- Now we can make word cloud. For creating word cloud, we use this code-

```
wordcloud(words = names(w),
          freq = w,
          max.words = 300,
          random.order = FALSE,
          min.freq = 50,
          colors = brewer.pal(8,
                            'Dark2'),
          scale = c(5, 0.3))
```
- Here we use maximum 300 words in the word cloud and minimum frequency is 50 that means any word that appears 50 times will be included and here we specified color dark 2. We changed the scale as highest frequency word as 5 and lowest frequency word 0.3. Running this code,

we got a word cloud looks like this (Figure 5)-



Fig. 5. Word Cloud Creation in R Studio

Word Cloud 2

- Another interesting word cloud we can create through word cloud 2. So, we updated a library with word cloud 2 by this code: `library(wordcloud2)`
- Then we stored this data into data frame we have names stored in `w` and frequency is stored in `w` by using this code: `w <- data.frame(names(w), w)`

- Then we give column names for this w, The first column name is words and second column name is frequency by using this code: `colnames(w) <- c('word', 'freq')`
- For example, we use this code- `head(w)` that gives word and frequency column.
- With this we can make word cloud 2. So first we specify w in word cloud 2 then specify size and specify default shape which is circle by this code (Figure 6)-
`wordcloud2(w,`
`size = 0.8,`
`shape = 'circle')`

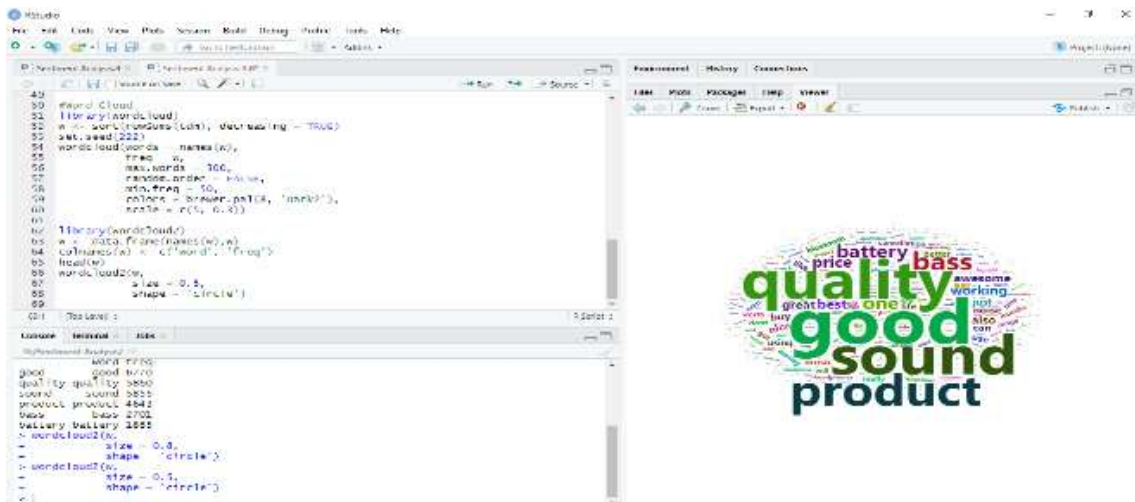


Fig. 6. Word Cloud 2 in Circle Shape

- Then we change the size again 0.8 to 0.5 to get a better look.
- Then we use another shape star and rotation ratio and minimum size to get another look and the code is looks like now (Figure 7)-

```
wordcloud2 (w,
  size = 0.5,
  shape = 'star',
  minSize = 1,
  rotateRatio = 0.8)
```

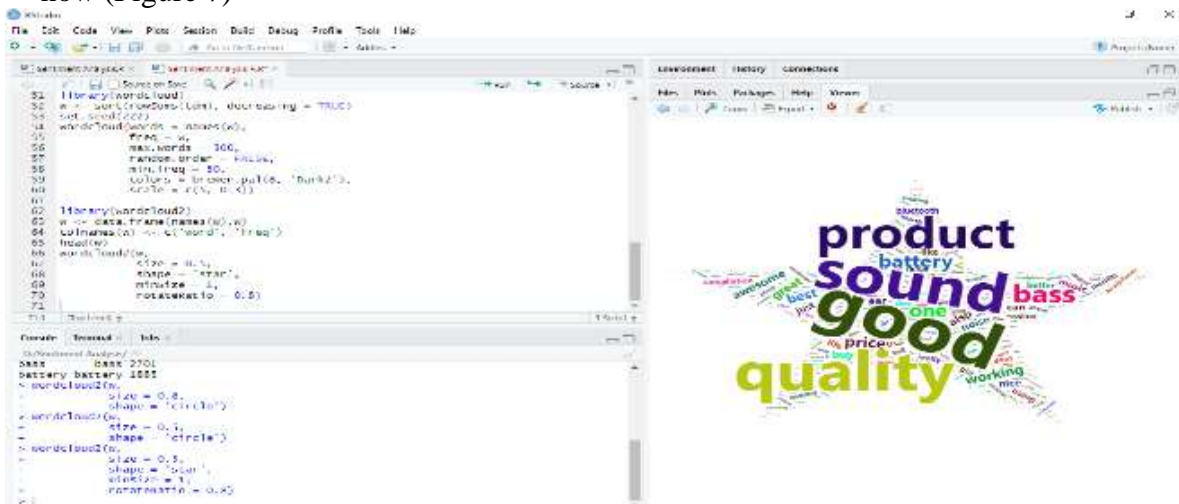


Fig. 7. Word Cloud 2 in Star Shape

- For triangle shape the code is (Figure 8):
`wordcloud2 (w,`
`size = 0.5,`
`shape = 'triangle',`
`minSize = 1,`
`rotateRatio = 0.8)`

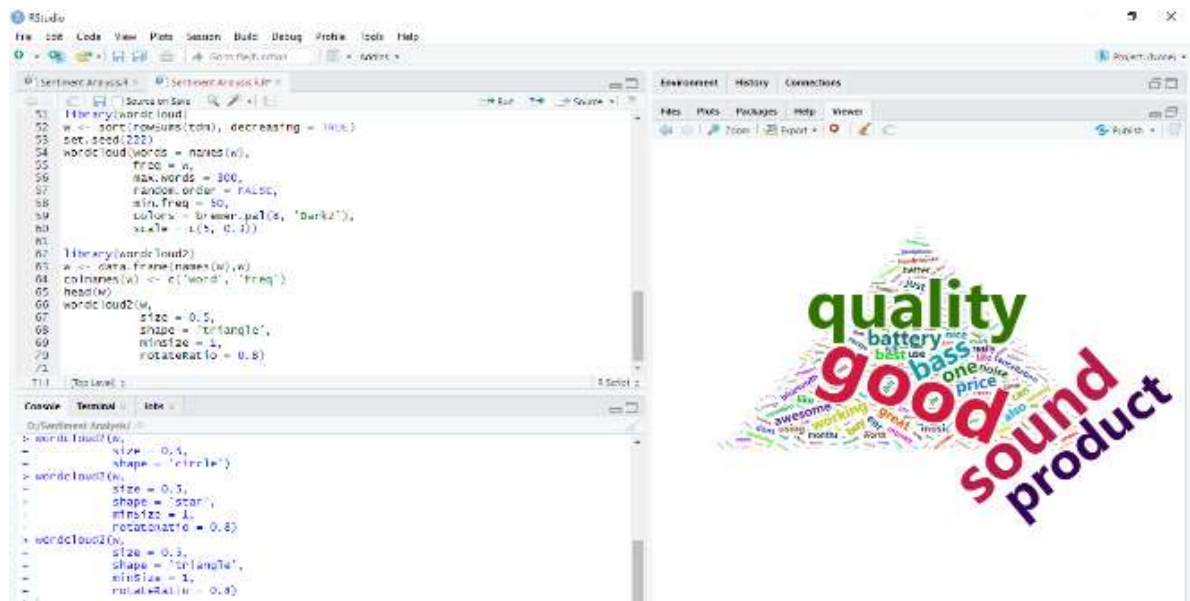


Fig. 8. Word Cloud 2 in Triangle Shape

Sentiment Analysis

For sentiment analysis of those reviews we need some packages, and those packages are: **syuzhet**, **lubridate**, **ggplot2**, **scales**, **reshape2** and **dplyr**.

- As we have not used them before so firstly, we need to install them by going to install packages. After installing them we simply ran the library (Figure 9).

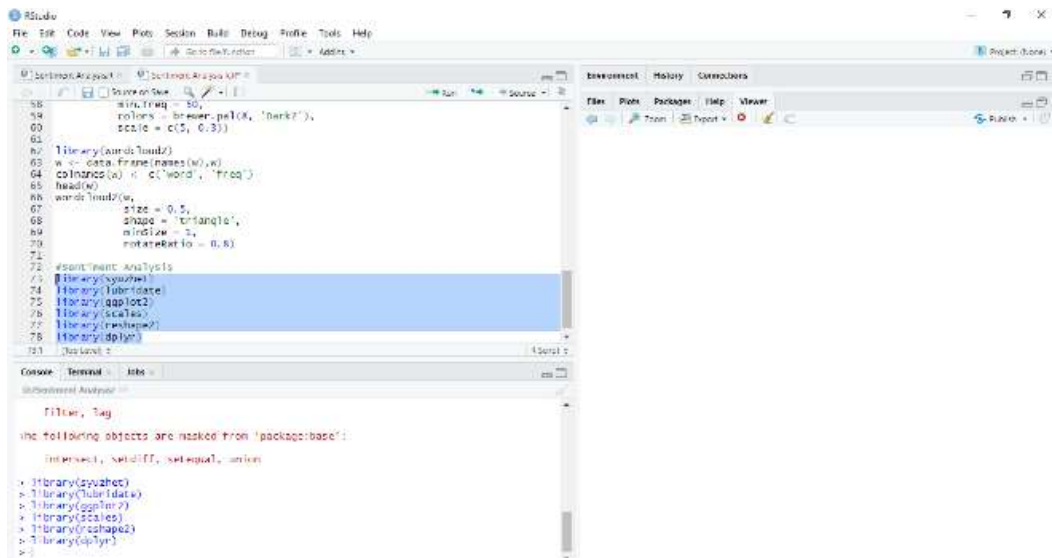


Fig. 9. Sentiment Library Installation

Read file

- We read the file with this code-
`earphones <- read.csv(file.choose(), header = T)` and stored it in earphones as our dataset is all about earphones.
- Then we stored only **ReviewBody** column which is about reviews into review with this code-
`review <-`

```
iconv(earphones$ReviewBody, to = 'utf8')
```

Obtain Sentiment Score

- To get sentiment score (Figure 10) we used `get_nrc_sentiment` that means nrc sentiment dictionary. This dictionary will provide score for each and every word for different emotions. So, we use the code to

use nrc dictionary and stored it in s: `s <- get_nrc_sentiment(review)`

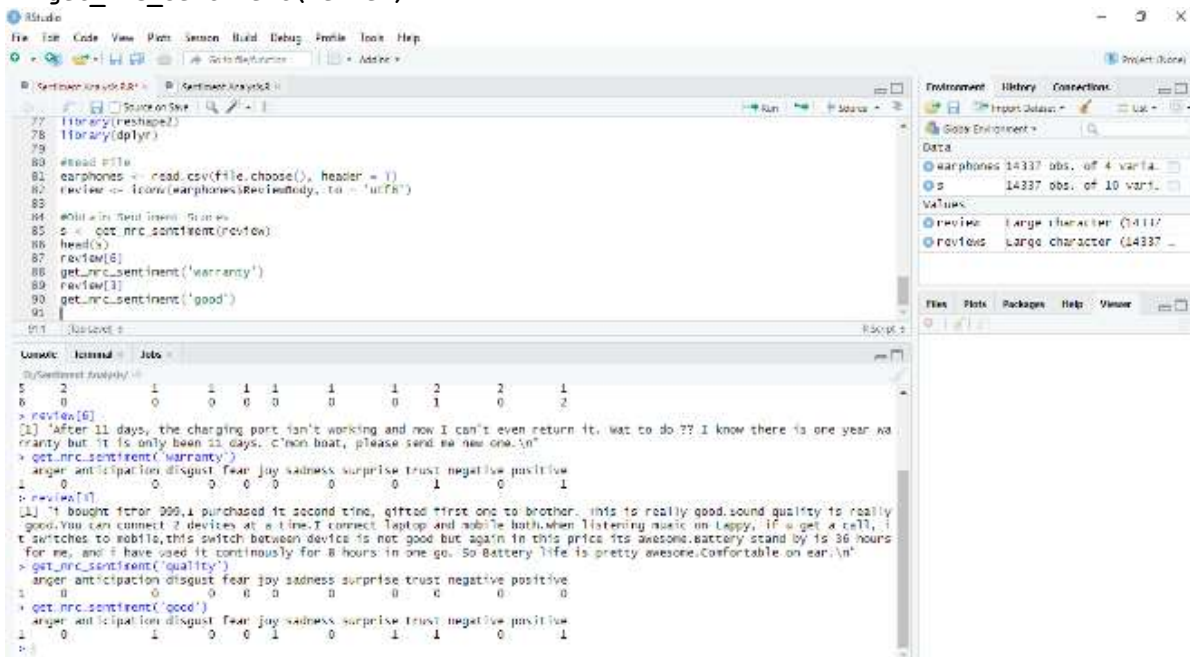


Fig. 10. Obtaining Sentiment Score for Specific Word

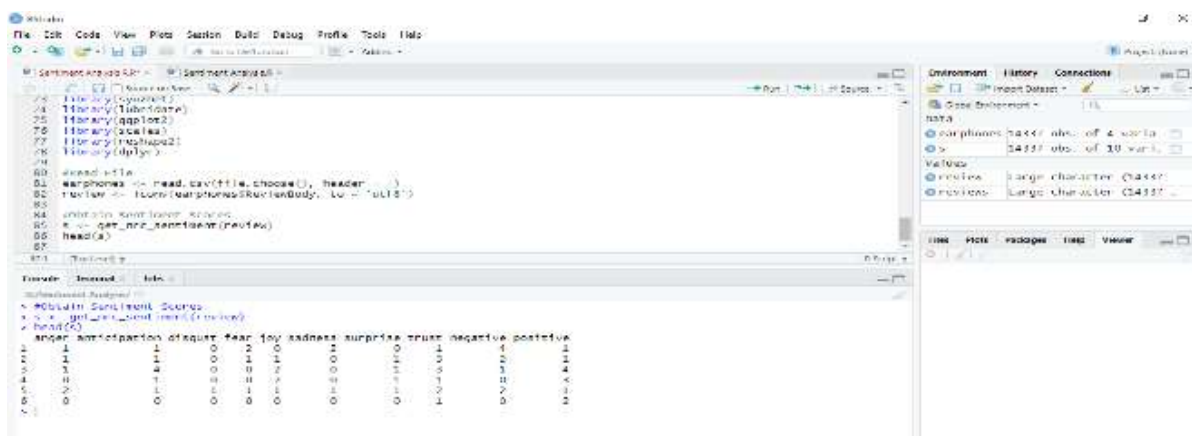


Fig. 11. Obtaining Sentiment Score

- As we want to know how it looks so we ran this code: `head(s)`, that shows 10 different columns (Figure 11) which are anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative and positive. It also shows scores for some reviews. For example, number six review have scores 1 for trust and 2 for positive.
- To get better understanding we tried to get score for a specific word 'warranty' from review number six. So, we use these code-

```
review [6]
get_nrc_sentiment('warranty')
```

This shows that warranty has scored 1 for trust and 1 for positive.
- Again, we checked two words 'quality' and 'good' from review number 3 in the same way and it shows that quality doesn't have any score and good has scored 1 in anticipation, joy, surprise, trust and positive.

Bar Plot

- We can get a bar plot (Figure 12) which will represent the sentiment of all reviews by using this code-

```
#Barplot
barplot(colSums(s),
        las=2,
        col = rainbow(10),
        ylab = 'count',
```

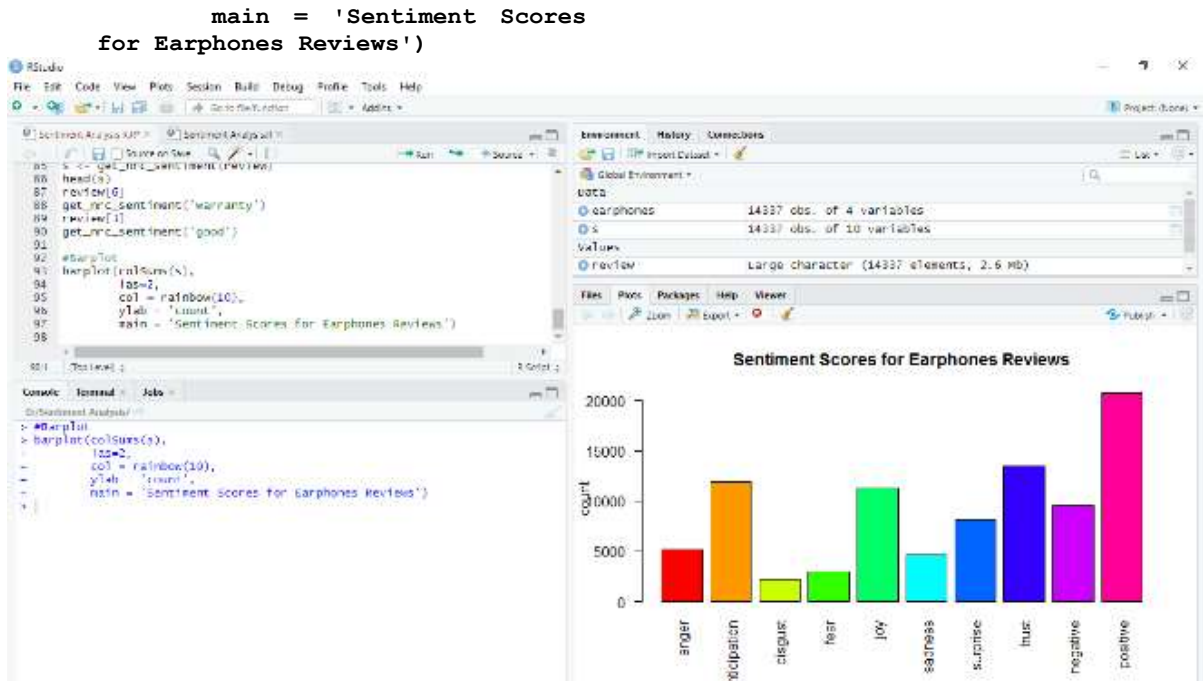



Fig. 12. Bar Plot for Sentiment Analysis

Here we titled the bar plot as ‘Sentiment Scores for Earphones Reviews’, used color `rainbow(10)` as there are only ten columns. we used y axis for count (Figure 13).

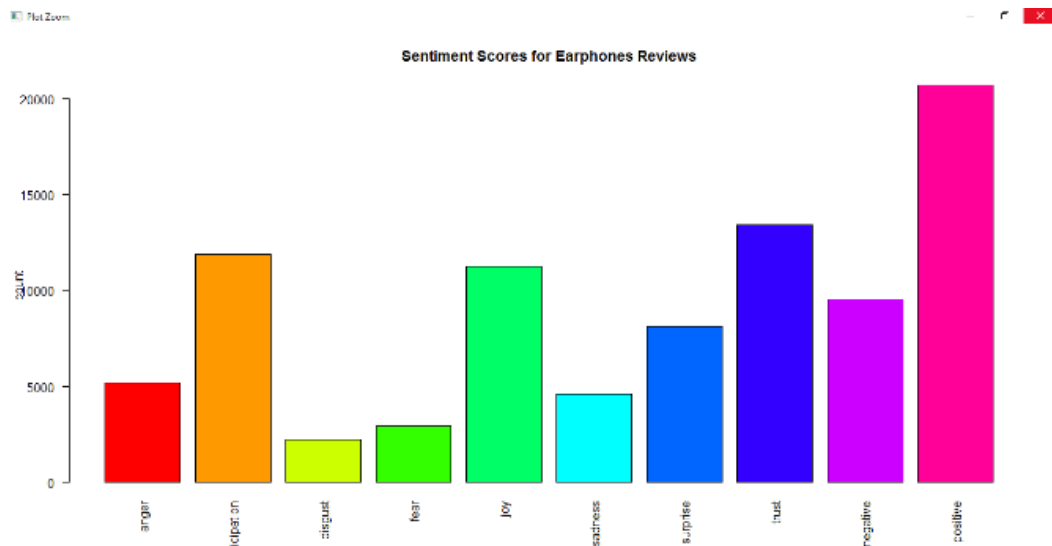


Fig. 13. Sentiment Score for Amazon Earphones

4 Research Findings

From this bar plot we can see that highest bar is positive. So, it can be said that sentiment for those earphones is positive. That means people like those earphones. It also shows that trust, joy, anticipation and surprise is quite high rather than negative, sadness, fear, disgust and anger. It means those brands are really doing well in the market.

This is how we can get sentiment from users before and after any incident and it will help to take further decision.

5 Conclusion

From this paper our main aim was to learn the process of making word cloud and sentiment analysis in R language. At the last, it can be said that our aim or goal is achieved because word cloud is made and the sentiment analysis

is also done for amazon earphones reviews which helps those producers to take decision for further marketing those products or to develop the product or to introduce new product in the market in a customer centric way. we extracted here about 14 thousand unstructured reviews and come up with word cloud and attitude of customers towards those products. So, our main goal is achieved here.

References

- [1] I. Habernal, T. Ptáček, and J. Steinberger, "Sentiment analysis in Czech social media using supervised machine learning," in Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, 2013, pp. 65-74.
- [2] R. C. Team, "R language definition," Vienna, Austria: R foundation for statistical computing, 2000.
- [3] I. Feinerer, "Introduction to the tm Package Text Mining in R," Accessible en ligne: <http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>, 2013.
- [4] D. Krieger, "Corpus linguistics: What it is and how it can be applied to teaching," The Internet TESL Journal, vol. 9, 2003.
- [5] E. De Jonge and M. Van Der Loo, An introduction to data cleaning with R: Statistics Netherlands Heerlen, 2013.
- [6] C. Facer. (2017, February 27). Text Analysis: Hooking up Your Term Document Matrix to Custom R Code. Available: <https://www.displayr.com/text-analysis-hooking-up-your-term-document-matrix-to-custom-r-code/>
- [7] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1833-1842.
- [8] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends® in Information Retrieval, vol. 2, pp. 1-135, 2008.
- [9] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the Workshop on Language in Social Media (LSM 2011), 2011, pp. 30-38.
- [10] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, pp. 1-167, 2012.
- [11] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R," Informatica Economica, vol. 22, 2018.
- [12] J. S. Brown and P. Duguid, The social life of information: Updated, with a new preface: Harvard Business Review Press, 2017.
- [13] P. Casanovas, L. De Koker, D. Mendelson, and D. Watts, "Regulation of Big Data: Perspectives on strategy, policy, law and privacy," Health and Technology, vol. 7, pp. 335-349, 2017.
- [14] H. Kennedy, "Perspectives on sentiment analysis," Journal of Broadcasting & Electronic Media, vol. 56, pp. 435-450, 2012.
- [15] J. Feine, S. Morana, and U. Gnewuch, "Measuring service encounter satisfaction with customer service chatbots using sentiment analysis," 2019.
- [16] J. J. Hox and H. R. Boeijs, "Data collection, primary versus secondary," 2005.
- [17] J. Chatman, C. O'Reilly, and V. Chang, "Cisco Systems: Developing a human capital strategy," California Management Review, vol. 47, pp. 137-167, 2005.



Ahmed Imran Kabir has graduated from the college of Business with a Master of Science in Business Analytics/ Data Science from Texas A&M University- Commerce, United States in 2017. He is now a doctoral candidate in the Department of Business and Accountancy, University of Malaya, Malaysia. His work focuses on healthcare analytics, business intelligence, text mining, and big data software.



Ridoan Karim has graduated from the Department of Law with a Master of Comparative Law from International Islamic University Malaysia (IIUM) in 2016. He is now a doctoral candidate in the Department of Law; University of Malaya, Malaysia. His work focuses of nuclear regulations, data protection law, block chain technology. He worked as an adjunct lecturer at East Delta University, Bangladesh in the Fall semester 2016.

Appendix – Source Code

```
#Read File
> earphones <- read.csv(file.choose(), header = T)
> str(earphones)

#Build Corpus
> library[17]
> corpus <- iconv(earphones$ReviewBody, to = 'utf8')
> corpus <- Corpus(VectorSource(corpus))
> inspect(corpus[1:5])

#Clean Text
> corpus <- tm_map(corpus, tolower)
> inspect(corpus[1:5])

> corpus <- tm_map(corpus, removePunctuation)
> inspect(corpus[1:5])

> corpus <- tm_map(corpus, removeNumbers)
> inspect(corpus[1:5])

> cleanset <- tm_map(corpus, removeWords, stopwords('english'))
> inspect(cleanset[1:5])

> removeLast <- function(x) gsub('\n', '', x)
> cleanset <- tm_map(cleanset, content_transformer(removeLast))
> inspect(cleanset[1:5])

> removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)
> cleanset <- tm_map(cleanset, content_transformer(removeURL))
> binspect(cleanset[1:5])

> cleanset <- tm_map(cleanset, removeWords, c('earphone', 'earphones',
'boat'))

> cleanset <- tm_map(cleanset, stripWhitespace)
> inspect(cleanset[1:5])

#Term Document Matrix
> tdm <- TermDocumentMatrix(cleanset)
> tdm
> tdm <- as.matrix(tdm)
> tdm[1:10, 1:20]

#Bar Plot
> w <- rowSums(tdm)
> w <- subset(w, w>=300)
> barplot(w,
         las=2,
         col = rainbow(50))

#Word Cloud
> library(wordcloud)
> w <- sort(rowSums(tdm), decreasing = TRUE)
> set.seed(222)
> wordcloud(words = names(w),
           freq = w,
           max.words = 300,
           random.order = FALSE,
```



```
      min.freq = 50,
      colors = brewer.pal(8, 'Dark2'),
      scale = c(5, 0.3))

> library(wordcloud2)
> w <- data.frame(names(w),w)
> colnames(w) <- c('word', 'freq')
> head(w)
> wordcloud2(w,
             size = 0.5,
             shape = 'triangle',
             minSize = 1,
             rotateRatio = 0.8)

#Sentiment Analysis
library(syuzhet)
library(lubridate)
library(ggplot2)
library(scales)
library(reshape2)
library(dplyr)

#Read File
> earphones <- read.csv(file.choose(), header = T)
> review <- iconv(earphones$ReviewBody, to = 'utf8')

#Obtain Sentiment Scores
> s <- get_nrc_sentiment(review)
> head(s)
> review[6]
> get_nrc_sentiment('warranty')
> review[3]
> get_nrc_sentiment('good')

#Barplot
> barplot(colSums(s),
         las=2,
         col = rainbow(10),
         ylab = 'count',
         main = 'Sentiment Scores for Earphones Reviews')
```