

**LAPORAN PROYEK AKHIR  
PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN TERHADAP REVIEW  
PRODUK EARPHONE OLEH PELANGGAN  
E-COMMERCE AMAZON**



Aliefa Haniif Setyaningrum 123190090  
Azzahra Galih Ayudya 123190104

**PROGRAM STUDI INFORMATIKA  
JURUSAN INFORMATIKA  
FAKULTAS TEKNIK INDUSTRI  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
YOGYAKARTA  
2021**

## 1. PENDAHULUAN

Perkembangan teknologi saat ini sangat pesat dan segalanya dilakukan secara digital. Tidak hanya dalam bidang informasi dan komunikasi namun bidang lainnya seperti bisnis hingga berbelanja bergantung dengan teknologi. Saat ini berbelanja bisa dilakukan tanpa harus menuju ke tempat penjual atau toko namun bisa dilakukan secara online melalui aplikasi e-commerce. banyak kebutuhan yang dapat dibeli berdasarkan review yang ada di produk itu.

Amazon merupakan toko online (*e-commerce*) yang bergerak dibidang penjualan berbagai produk baik produk rumah tangga, elektronik, pertukangan, sparepart kendaraan dan masih banyak lagi. Pelanggan dapat membeli produk melalui aplikasi Amazon dengan cara mencari produk yang diinginkan di handphone pelanggan. Biasanya sebelum membeli barang di e-commerce, pelanggan akan melihat reviews produk dibagian bawah produk yang dijual. Review ini berisi ulasan atau review pelanggan terhadap suatu produk. Review dari pelanggan ini dapat digunakan sebagai tolak ukur yang efektif untuk melihat keaslian dan informasi lain dari produk apakah sesuai dengan yang dicantumkan atau tidak.

Sentiment analysis atau analisis sentimen dalam bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif. Hasil sistem prototipe mencapai tinggi presisi (75-95% tergantung pada data) dalam mencari sentimen pada halaman web dan artikel berita. (Nasukawa.T.& Yi, 2003). Analisis sentimen sangat berguna untuk menganalisis review yang diberikan pelanggan dan kemudian diproses menjadi sesuatu yang memiliki nilai atau informasi.

Misalkan dalam project ini, kita akan melakukan analisis sentimen terhadap produk earphone kepada pelanggan Amazon berdasarkan review yang diberikan oleh pelanggan yang lain. Semakin berkembangnya teknologi, earphone sekarang ini memiliki banyak spesifikasi dan merek tertentu yang terkadang membuat orang susah dalam memilih earphone mana yang sesuai dengan yang diinginkan dan dibutuhkan, dan dengan spesifikasi yang seperti apa atau dengan harga mana yang paling

terjangkau. Dengan masalah itu, pelanggan akan membutuhkan suatu sentimen dimana pelanggan dapat menilai sendiri apakah produk earphone yang dijual tersebut memiliki review yang positif atau negatif.

## **2. METODE**

Metode dalam penelitian ini terdiri dari *data collection*, *data cleaning*, dan *data processing* menggunakan *naive bayes*. Ketiga tahap tersebut akan dilakukan menggunakan RStudio yaitu software yang digunakan untuk menulis code menggunakan bahasa R. RStudio merupakan *integrated development environment* (IDE) yang didalamnya terdapat konsol, editor penyorotan sintaks yang mendukung eksekusi kode langsung, serta alat untuk merencanakan, riwayat, debugging, dan manajemen ruang kerja. Bahasa R merupakan salah satu bahasa pemrograman yang digunakan untuk memanipulasi, mengkalkulasi, dan menganalisa data. R disebut bahasa pemrograman karena dapat melakukan conditional, looping, dan berbagai macam fungsi input output lainnya.

### **3.1 Data Collection**

Dalam penelitian ini, data yang akan dianalisa adalah data yang diperoleh dari Kaggle dengan judul Amazon Earphone Reviews. Data ini terdiri dari 14.337 reviews yang dilakukan oleh pelanggan Amazon. Pengambilan datanya diperoleh dari hasil ekstrak yang dilakukan pada website Amazon.

### **3.2 Data Cleaning**

Data cleaning atau pembersihan data dilakukan untuk mendapatkan data-data tertentu agar dapat dianalisa. Data cleaning diproses menggunakan bahasa R. Tahapannya yaitu :

- 1) Data pertama kali akan dicek tipenya dan dinormalizing sehingga data dapat diproses secara teknis. Proses ini terdiri dari menghapus URL, hyperlinks, HTML tags, subjek, kata ganti, spasi, tanda baca, dan karakter sejenis lainnya.
- 2) Memperbaiki dan menghubungkan isi dari data yang sudah dinormalisasi. Data yang sudah dinormalisasi masih memiliki kemungkinan error sehingga

harus diperbaiki atau dipilah sesuai dengan kebutuhan untuk menghilangkan ketidakkonsekuensi.

- 3) Data yang sudah konsisten tersebut kemudian dilakukan analisa sesuai kebutuhan.

### **3.3 Data Processing menggunakan *Naïve Bayes Classifier***

*Naïve Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema *Bayes* . Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi/kejadian.

Kegunaan *Naïve Bayes* :

- Sebagai metode machine learning yang menggunakan probabilitas
- Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis
- Untuk membuat diagnosis medis secara otomatis
- Mendeteksi atau menyaring spam

Kelebihan *Naïve Bayes*

- Bisa dipakai untuk data kuantitatif maupun kualitatif
- Tidak memerlukan jumlah data yang banyak
- Tidak perlu melakukan data training yang banyak
- Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
- Perhitungannya cepat dan efisien
- Mudah dipahami
- Mudah dibuat

- Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang
- Jika digunakan dalam bahasa pemrograman, *code*-nya sederhana
- Bisa digunakan untuk klasifikasi masalah biner ataupun *multiclass*

### Kekurangan Naïve Bayes

- Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
- Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
- Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya.
- Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut Banyak celah yang bisa mengurangi efektivitasnya  
Dirancang untuk mendeteksi kata-kata saja, tidak bisa berupa gambar

Dalam project ini, dilakukan pengklasifikasian terhadap 14.337 reviews oleh pengguna e-commerce Amazon terhadap produk earphone. Dengan metode *Naïve Bayes Classifier*, review ini dibagi dibagi menjadi dua kategori kelas yaitu review positif dan review negatif.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Daftar Ulasan Produk Earphone

Data diperoleh dari [www.kaggle.com](http://www.kaggle.com). Review yang digunakan didapatkan dari kegiatan pembelian produk Bluetooth Earphone di Amazon. Ulasan yang digunakan sebanyak 14, 337 ulasan. Dataset ini menggunakan ulasan berbahasa Inggris. Berikut adalah dataset dalam bentuk datasheet.

ReviewTitle	ReviewBody	ReviewStar	Product
1 Honest review of an edm music lover	No doubt it has a great bass and to a great extent noise cancellation and decent sound clarity and mindblowing battery but the following dissapointed me though I tried a lot to adjust.1 Bluetooth range not more than 10m2. Pain in ear due the conical buds(can be removed)3. Wires are a bit long which makes it odd in front 4. No pouch provided 5. Worst part is very low quality and distoring mic. Other person keeps complaining about my voice.	3	boAt Rockerz 255
2 Unreliable earphones with high cost	This earphones are unreliable. I bought it before 15 days meanwhile right side ear buds got cracked automatically and it got divided in two parts, and sound quality is also not that much good but ok, one more thing bass is not good as it is a boat earphones Guys,Also for the proof i have attached picsPlease see and think before buying?? this unreliable product.Thanks.	1	boAt Rockerz 255
3 Really good and durable.	I bought itfor 999.I purchased it second time, gifted first one to brother. This is really good.Sound quality is really good.You can connect 2 devices at a time.I connect laptop and mobile both.when listening music on Lappy, if u get a call, it switches to mobile.this switch between device is not good but again in this price its awesome.Battery stand by is 36 hours for me, and i have used it continuously for 8 hours in one go. So Battery life is pretty awesome.Comfortable on ear.	4	boAt Rockerz 255
4 stopped working in just 14 days	Its sound quality is adorable. overall it was good but just for 2 weeks after that it stopped working and since then it never powered on again. I am claiming for warrenty and it is still undergoing.	1	boAt Rockerz 255
5 Just Awesome Wireless Headphone under 1000. 0777	Its Awesome. Good sound quality & 8-9 hrs battery life. Just waww look. 0777 I brought just 999 rupees. But Amazon's Packageing system is Very bad. Very poor packaging. Flipcart's packaging system is much better than Amazon.	5	boAt Rockerz 255
6 Charging port not working	After 11 days, the charging port isn't working and now I can't even return it. Wat to do ?? I know there is one year warranty but it is only been 11 days. C'mon boat, please send me new one.	1	boAt Rockerz 255
7 Loved the color and sound !!	Sound-4.5/5Build quality-4.5/5Battery life -4/5Earphones quality is good with strong magnetic lock. Straps is good quality and the earphones wires are too good. Buttons quality is good. sound is full of punchy bass and and balanced treble and has somund effect. good fitting in ears. hope connectivity is good. bcz my last boat Bluetooth bassheads 225 had connection issues. This is worth its price. Hope the packaging was good. there was no bubble wrapping.	4	boAt Rockerz 255
8 Great sound with worst mic.	After one month usage reviewing this Pros-1- sound quality2- design3- weight4- batteryCons1- very bad mic for talking2- can't use while riding bike3 - Music is too loud	3	boAt Rockerz 255
9 Battery life is good enough to keep going all day	I can't get enough of this earphone! The bass is amazing and is perfect for a music lover. The battery life is good enough to keep me going all day. The fact that I can connect my laptop and phone at the same time and switch between them makes the experience so much better. I use my earphones a lot for voice calls and this doesn't let me down at all. This is a must buy for all music lovers.	5	boAt Rockerz 255
10 Very very bad Durability	The product durability is 1 month. I just lost my bill details so I cant claim my warrenty. But this earphones built quality is bad. It stops wrking completely after 1 month sep 14 to nov 11 0777 life span. Dont lose ur bill details incase u want that warranty. And after purchasing u must register ur product in boat's site for replacement and warranty	1	boAt Rockerz 255

Gambar 3.1.1 Dataset

Dataset kemudian dilakukan tahap pra-pemrosesan diterapkan fitur *word normalizer*, *stemming*, dan *stopword removal* untuk setiap ulasan. *Word normalizer* untuk menangani variasi penulisan kata yang memiliki makna yang sama agar terhitung sebagai istilah tunggal. Penggunaan word normalizer dapat menangani banyaknya variasi kata yang digunakan pada penulisan ulasan oleh pembeli untuk diubah menjadi istilah yang sama. Selanjutnya dilakukan proses *stopword removal* untuk menghilangkan stopwords dari ulasan. Daftar stopwords yang digunakan dibuat sendiri mengacu pada konteks kata yang sering dipakai dalam ulasan maupun dalam jual beli secara daring. Selain dalam bentuk kata angka juga masuk ke dalam stopwords, angka tidak berpengaruh pada analisis sentimen dan dapat dihapus, sehingga dapat mengurangi kebisingan dan meningkatkan efisiensi. Berikut dataset hasil dari pra-pemrosesan.

	text	sentimen
1	doubt great bass great extent noise cancellation decent sound clarity mindblowing battery following dissapointed though tried lot adjustbluetooth range m pain ear due conical buds can removed wires bit long makes odd front pouch provided worst part low quality distoring mic person keeps complaining voice	Negative
2	earphones unreliable bought days meanwhile right side ear buds got cracked automatically got divided two parts sound quality also much good ok one thing bass good boat earphonesguysalso proof attached picplease see think buying unreliable productthanks	Negative
3	bought itfor purchased second time gifted first one brother really goodsound quality really goodyou connect devices timei connect laptop mobile bothwhen listening muaic lappy u get call switches mobilethis switch device good price awesomebattery stand hours used continuously hours one go battery life pretty awesomecomfortable ear	Negative
4	sound quality adorable overall good weeks stopped working since never powered claiming warrenty still undergoing	Positive
5	awesome good sound quality hrs battery life waww look brought rupeesbut amazons packageing system bad poor packaging flipcarts packaging system much better amazon	Negative
6	days charging port isnt working cant even return wat know one year warranty days cmon boat please send new one	Positive
7	soundbuild qualitybattery life earphones quality good strong magnetic lock straps good quality earphones wires good buttons quality good sound full punchy bass balanced treble sorund effect good fitting ears hope connectivity good bcz last boat bluetooth bassheads connection issues worth price hope packaging good bubble wrapping	Positive
8	one month usage reviewing pros sound quality design weight batterycons bad mic talking cant use riding bike music loud	Neutral
9	cant get enough earphone bass amazing perfect music lover battery life good enough keep going day fact connect laptop phone time switch makes experience much better use earphones lot voice calls doesnt let must buy music lovers	Negative
10	product durability monthi lost bill details cant claim warrenty earphones built quality bad stops wrking completely month sep nov life span dont lose ur bill details incase u want warranty purchasing u must register ur product boats site replacement warranty	Negative

Gambar 3.1.2 Dataset Hasil Pra-Pemrosesan

## 3.2 Pembahasan

```
library(tm) #membersihkan data
library(vroom) #load dataset
library(here) #menyimpan dataset

d <- vroom(here('AllProductReviews.csv'))
review1 <- d$ReviewBody
review2<- Corpus(VectorSource(review1))
removeURL <- function(x) gsub("http[^\[:space:]]*", "", x)
reviewclean <- tm_map(review2, removeURL)
removeNL <- function(y) gsub("\n", " ", y)
reviewclean <- tm_map(reviewclean, removeNL)
replacecomma <- function(y) gsub(",", "", y)
reviewclean <- tm_map(reviewclean, replacecomma)
removetitik2 <- function(y) gsub(":", "", y)
reviewclean <- tm_map(reviewclean, removetitik2)
removetitikkoma <- function(y) gsub(";", " ", y)
reviewclean <- tm_map(reviewclean, removetitikkoma)
removetitik3 <- function(y) gsub("p..", "", y)
reviewclean <- tm_map(reviewclean, removetitik3)
removeamp <- function(y) gsub("&", "", y)
reviewclean <- tm_map(reviewclean, removeamp)
```

```

removeUN <- function(z) gsub("@\\w+", "", z)

reviewclean <- tm_map(reviewclean, removeUN)

remove.all <- function(xy) gsub("[^[:alpha:][:space:]]*", "",
xy)

reviewclean <- tm_map(reviewclean,remove.all)

reviewclean <- tm_map(reviewclean, removePunctuation)

reviewclean <- tm_map(reviewclean, tolower)

hapus = readLines("hapuskata.txt")

reviewclean <- tm_map(reviewclean,removeWords,hapus)


dataframe<-data.frame(text=unlist(sapply(reviewclean, `[]`)),
stringsAsFactors=F)

View(dataframe)

write.csv(dataframe,file = 'clean_earphone.csv')

```

Program diatas digunakan untuk membersihkan dataset. Proses ini terdiri dari menghapus URL, hyperlinks, HTML tags, subjek, kata ganti, spasi, tanda baca, dan karakter sejenis lainnya.

```

library(e1071) #untuk naive bayes
library(caret) #untuk klasifikasi data
library(syuzhet) #untuk membaca fungsi get_nrc

datareview <- read.csv("clean_earphone.csv",stringsAsFactors =
FALSE)

review <- as.character(datareview$text) #merubah text menjadi
char

s <- get_nrc_sentiment(review)

review_combine <- cbind(datareview$text,s) #klasifikasi data

#Pembentukan hasil sentimen analisis NRC

sentiment <- data.frame(negative=s$negative,
positive=s$positive)

classify <- mutate(sentiment, text_sentiment=
ifelse((sentiment$negative != sentiment$positive),

```



```

    ifelse(sentiment$negative !=0, print("Negative"),

    print("Positive")),

    print("Neutral")

    ))

df<-data.frame(text=review,sentimen=classify$text_sentiment)

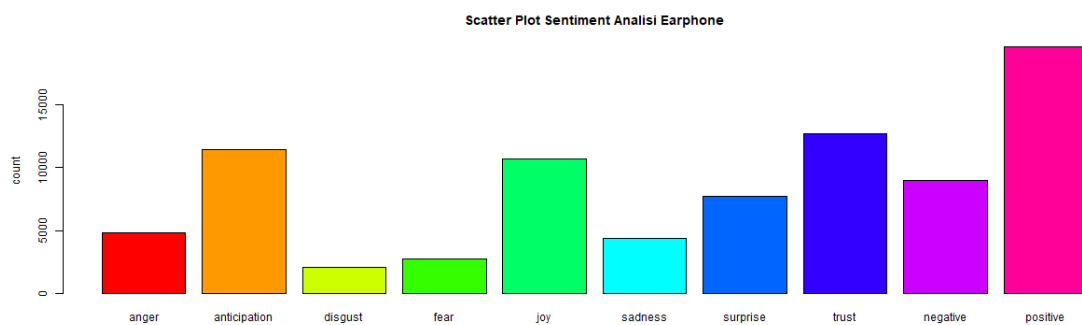
par(mar=rep(3,4))

a <-
barplot(colSums(s),col=rainbow(10),ylab='count',main='Scatter
Plot Sentiment Analisi Earphone')

brplt <- a

```

Program di atas untuk mengimport dataset. Selain itu dilakukan pembentukan sentimen analisis berupa output positif atau negatif. Kemudian dari data yang sudah diimport dihasilkan pula plot berupa Scatterplot.



Gambar 3.1.3 Scatter plot

```

library(tm)

data_earphone=as.factor(datareview$text)

corpus= Corpus(VectorSource(data_earphone))

fw <- TermDocumentMatrix(corpus)

fw <- removeSparseTerms(fw, sparse=0.98)

fw <- as.matrix(fw)

w = sort(rowSums(fw),decreasing = TRUE)

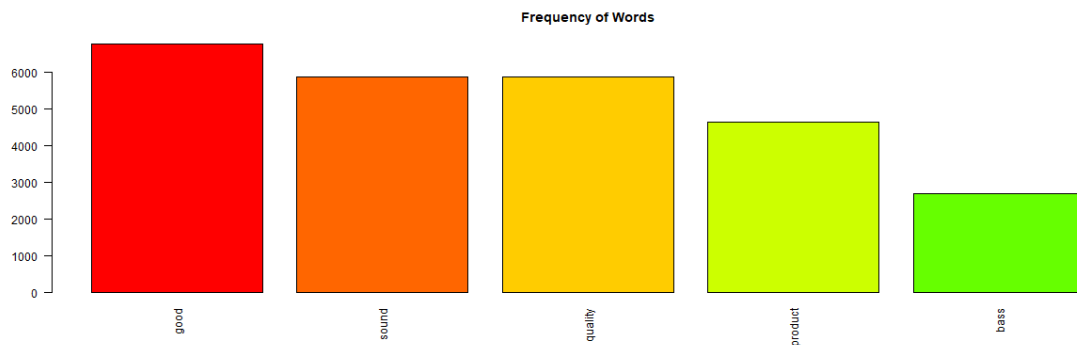
```

```

barplot(w[1:5],
        las=2,
        main="Frequency of Words",
        col= rainbow(15))

```

Program diatas untuk menampilkan plot berupa Frekuensi Plot dimana akan memberikan keterangan berapa frekuensi suatu kata dalam dataset.



Gambar 3.1.4 Frekuensi Plot

```

#library untuk penggunaan corpus dalam cleaning data

library(tm)

library(RTextTools)

#library yang terdapat sebuah algoritma naivebayes

library(e1071)

library(dplyr)

library(caret)

df<-read.csv("clean_earphone.csv",stringsAsFactors = FALSE)

glimpse(df)

#Set the seed of R's random number generator, which is useful
for creating simulations or random objects that can be
reproduced.

set.seed(20)

df<-df[sample(nrow(df)),]

df<-df[sample(nrow(df)),]

glimpse(df)

```

```

corpus<-Corpus(VectorSource(df$text))
corpus
inspect(corpus[1:10])
#fungsinya untuk membersihkan data data yang tidak dibutuhkan
corpus.clean<-corpus%>%
  tm_map(content_transformer(tolower))%>%
  tm_map(removePunctuation)%>%
  tm_map(removeNumbers)%>%
  tm_map(removeWords, stopwords(kind="en"))%>%
  tm_map(stripWhitespace)
dtm<-DocumentTermMatrix(corpus.clean)

inspect(dtm[1:10,1:20])

df.train<-df[1:50,]
df.test<-df[51:100,]

dtm.train<-dtm[1:50,]
dtm.test<-dtm[51:100,]

corpus.clean.train<-corpus.clean[1:50]
corpus.clean.test<-corpus.clean[51:100]

dim(dtm.train)
fivefreq<-findFreqTerms(dtm.train, 5)
length(fivefreq)

dtm.train.nb<-DocumentTermMatrix(corpus.clean.train, control =
list(dictionary=fivefreq))

dtm.test.nb<-DocumentTermMatrix(corpus.clean.test, control =
list(dictionary=fivefreq))

```



```

        #tabPanel("Data Testing",DT::dataTableOutput('test')),
        tabPanel("Scatterplot",plotOutput('scatterplot')),
        tabPanel("Frequency Words", plotOutput('freq')),
        tabPanel("Worldcloud", plotOutput('wc'))
    )
)

#r server
library(shiny)
server <- function(input, output){

  output$data <- DT:: renderDataTable({
    DT::datatable(df, options = list(lenghtChange = FALSE))
  })

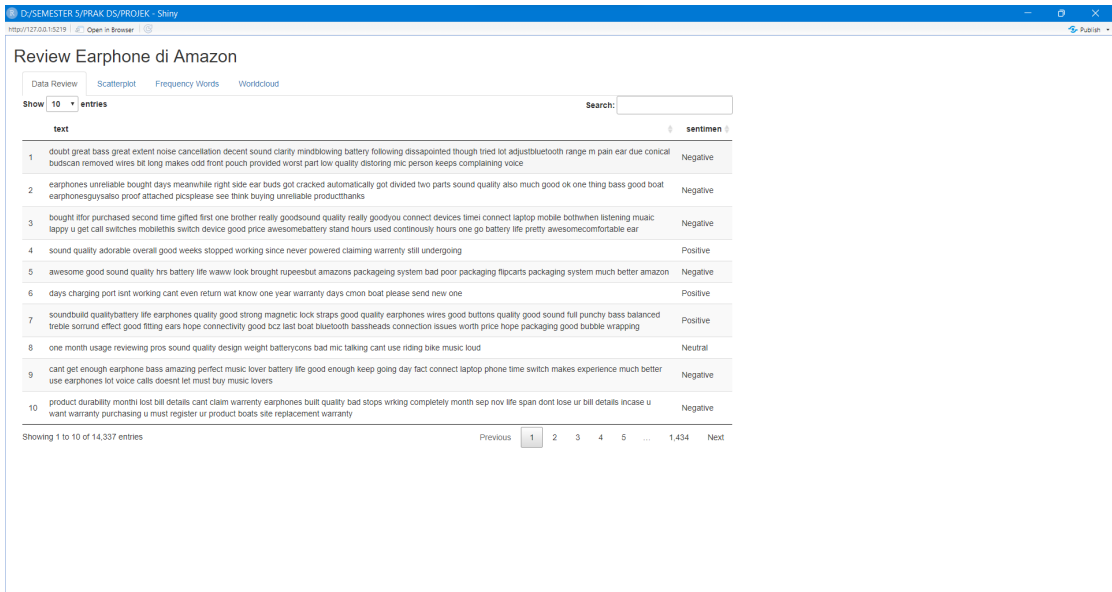
  output$scatterplot <- renderPlot({
    barplot(colSums(s),
            col=rainbow(10),
            ylab ='count',
            main='Scatter Plot Sentiment Analisi Earphone')
  })

  output$freq <- renderPlot({
    barplot(w[1:5],
            las=2,
            main="Frequency of Words",
            col= rainbow(15))
  })

  output$wc <- renderPlot({
    world_cloud<-wordcloud(corpus.clean,min.freq =
4,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))
  })
}
shinyApp(ui=ui,server=server,options=list(height="1080px"))

```

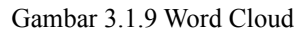
Di bawah ini merupakan gambar UI dari proyek yang dikerjakan :



Gambar 3.1.6 Data Review



Gambar 3.1.7 Scatter Plot



#### 4. KESIMPULAN

Berdasarkan project analisis sentimen terhadap review produk earphone oleh pelanggan e-commerce Amazon yang telah dilakukan menggunakan metode *Naïve Bayes Classifier*, dari 14.337 data review didapatkan hasil sebagai berikut :

##### 1) Scatter Plot

Dengan menggunakan Scatter Plot didapatkan hasil analisis sentimen menggunakan metode *Naïve Bayes Classifier* berdasarkan kategori review positif dan negatif. Review berkategori positif lebih banyak dibandingkan review berkategori negatif. Analisis sentimen berkategori positif dapat ditandai dengan "*trust*", "*surprise*", dan "*joy*". Sedangkan analisis sentimen berkategori negatif dapat ditandai dengan "*sadness*", "*fear*", dan "*disgust*".

##### 2) Frequency Plot

Dengan menggunakan Frequency Plot didapatkan hasil analisis kata yang sering digunakan pelanggan untuk menulis review berdasarkan posisinya dari frequency plot, kata yang paling sering digunakan berada di sebelah kiri yaitu "*good*" lalu diikuti dengan kata yang sering digunakan setelahnya "*sound*", "*quality*", "*product*", dan "*bass*".

##### 3) Word Cloud Plot

Dengan menggunakan Word Cloud Plot didapatkan gambar yang menunjukkan daftar kata-kata yang digunakan dalam sebuah teks, umumnya semakin banyak kata yang digunakan semakin besar ukuran kata tersebut dalam gambar. Dari gambar kita bisa mengetahui kata yang paling sering digunakan adalah "*good*" kemudian disusul dengan kata setelahnya, yaitu "*sound*", "*quality*", "*product*", dan "*bass*".



## DAFTAR PUSTAKA

Kurniawan, A. (2016). Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Metode Collaborative Filtering. In *Seminar Nasional Teknologi Informasi dan Komunikasi* (Vol. 2016, pp. 18-19).

AF IDA, K. H. U. S. N. A. (2018). *ANALISIS SENTIMENT TERHADAP JASA EKSPEDISI PENGIRIMAN BARANG POS INDONESIA MELALUI FACEBOOK DENGAN METODE NAÏVE BAYES DAN LEXICON BASED* (Doctoral dissertation, Universitas Muhammadiyah Ponorogo).

W. N. Venables, D.M. Smith, and R Core Team, An Introduction to R: Notes on R: Programming Environment for Data Analysis and Graphics Version 3.4.3, 2017.

Haryani, C. A., Tohari, H., Marhamah, M., & Nurrahman, Y. A. (2018). Sentimen Analisis Kepuasan Pelanggan E-commerce Menggunakan Lexicon Classification dengan R. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.