

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
ANALISIS SENTIMEN TERHADAP REVIEW
PENGGUNA AMAZON ALEXA**



CERTA YUSTITIA 123190093
TSANIAH MUNFIDAH 123190095

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

1. PENDAHULUAN

Amazon Alexa merupakan salah satu produk digital yang sekarang ini sangat populer dikalangan masyarakat dunia dan tak sedikit yang sudah memiliki perangkat ini. Alexa merupakan asisten cerdas yang dikembangkan oleh Amazon dan Amazon Echo dimana perangkat ini dapat melakukan komunikasi dua arah. Pada dasarnya Amazon Alexa merupakan *voice asisten digital* berteknologi *Artificial intelegent* yang dapat membalas pertanyaan sederhana dan dapat melakukan berbagai tugas atau perintah yang diberikan. Banyak kegiatan-kegiatan sehari-hari yang dapat secara efisien kita kerjakan dengan bantuan Alexa. Pada fitur smart home alexa dapat membantu kita mulai dari menghidupkan lampu dan TV, mengunci pintu rumah. Pada fitur Productivity Alexa dapat membantu kita dalam menyetel alarm dan mengingatkan berbagai agenda kita. Dan masih banyak lagi Fitur seperti membantu melakukan checkout shopping, memutar music bahkan membantu memonitor anak dan lain-lainnya. Semua kegiatan itu dapat dilakukan hanya dengan perintah melalui suara kita. Tak heran jika produk ini banyak digemari dan dibeli pada situs-situs penjualan terutama Amazon.

Dengan populernya Alexa dan semakin banyaknya pengguna yang membeli perangkat ini, dirasa perlu adanya analisis sentimen terkait review yang diberikan oleh pengguna. Analisis sentimen ini ditujukan agar dapat diketahui review atau pendapat atau komentar terkait produk ini baik review positif maupun negatif sehingga pengembang atau perusahaan dapat mengetahui permasalahan yang ada dan dapat meningkatkan kualitas produk Amazon Alexa ini.

Oleh karena itu dilakukan penelitian analisis sentimen terkait review Amazon Alexa baik review positif dan negatif yang datanya didapatkan melalui dataset dan akan ditampilkan hasilnya dalam bentuk barplot. Kemudian pada penelitian ini juga akan ditampilkan hasil berupa kata-kata dari hasil review yang banyak ditulis atau sering muncul dan akan ditampilkan dalam bentuk wordcloud. Selain itu dalam penelitian ini ditampilkan juga secara keseluruhan hasil analisis sentimen baik barplot hingga wordcloud ke dalam tampilan shiny.

2. METODE

2.1 Pengambilan Data

Pada penelitian analisis sentimen ini, data yang digunakan merupakan dataset yang diambil atau didownload dari situs [Kaggle](https://www.kaggle.com/sid321axn/amazon-alexa-reviews) dimana dataset ini merupakan dataset terkait review pengguna Amazon Alexa. Data yang didapatkan sendiri memiliki format berupa tsv yang kemudian kami ubah formatnya menjadi format csv. Dataset Review Pengguna Amazon Alexa ini terdiri dari berbagai data yang terbagi menjadi 5 buah kolom yaitu rating, date, variation, verified_reviews dan feedback serta memiliki data kurang lebih sekitar 3000 baris data. Dimana data yang akan digunakan adalah data pada kolom verified_reviews. Dataset yang ada dapat diunduh melalui link berikut ini: <https://www.kaggle.com/sid321axn/amazon-alexa-reviews>

2.2 Preprocessing atau Pembersihan Data

Proses yang pertama kali dilakukan setelah didapatkannya data adalah dilakukan pembersihan data dengan berbagai indikator. Tujuan dari dilakukannya pembersihan data ini sendiri adalah untuk mengekstraksi data set yang tidak terstruktur yang dapat mengganggu proses analisis data.

a. Replacement

Pada tahap ini dilakukan replacement atau penggantian terhadap beberapa data yaitu seperti URL (http), kemudian berbagai tanda baca ke dalam bentuk spasi

b. Removing

Pada tahap ini dilakukan removing terhadap beberapa hal yaitu seperti new line, tanda koma, tanda titik dua, tanda titik tiga, amp, karakter, tanda baca dan angka. Selain itu juga dilakukan removing kata-kata umum atau konjungsi-konjungsi yang tidak memiliki makna yang sudah disediakan dalam file stopwords.txt.

c. Content Transformer

Pada tahap ini data yang sudah dibersihkan akan diubah kedalam kalimat dengan huruf kecil.

2.3 Klasifikasi dengan Naive Bayes

Data yang telah dibersihkan pada proses preprocessing dan telah disimpan ke dalam file Ulasan_Clean.csv selanjutnya akan dibaca dan dibuat variable yang berguna untuk menampung data hasil dari nilai review. Kemudian dilakukan pembobotan pada setiap kategori kata berdasar pada algoritma naive bayes yang selanjutnya data akan disimpan ke dalam sebuah data frame.

2.4 Barplot

Data yang telah diklasifikasikan selanjutnya akan ditampilkan dalam bentuk Barplot. Barplot atau yang sering kita tahu sebagai diagram batang disini akan menampilkan perbandingan data pada satu atau lebih variabel yang nilainya sesuai dengan hasil klasifikasi pada data reviews. Data yang tampil pada barplot ini akan menunjukkan data mana yang memiliki nilai paling tinggi hingga ke rendah.

2.5 Wordcloud

Data juga ditampilkan ke dalam bentuk wordcloud. Wordcloud atau yang dikenal sebagai kolase kata merupakan representasi visual dari kata kata yang sering muncul pada suatu data text. Dengan menggunakan wordcloud ini, gambaran frekuensi kata-kata yang ditampilkan dari data review menjadi menarik, jelas dan informatif. Semakin sering suatu kata digunakan maka kata tersebut akan ditampilkan dengan ukuran yang semakin besar juga. Sehingga nantinya pada tampilan kata yang paling sering digunakan akan muncul dengan ukuran paling besar

2.6 Shiny

Shiny adalah *package* pada bahasa R yang memudahkan penggunaanya untuk membuat web interaktif dengan R. Shiny membantu mengubah analisis data menjadi sebuah web yang interaktif tanpa harus memahami tentang HTML, CSS ataupun JavaScript.

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil dan pembahasan dari analisis sentimen terhadap review pengguna alexa amazon yang telah dilakukan.

3.1 Library Untuk Data Cleaning

Untuk listing program 3.1 berisikan berbagai *library* yang dibutuhkan pada analisis ini diantaranya adalah *library* tm yang digunakan untuk pembersihan data, kemudian *library* vroom untuk load dataset dan *library* here yang berguna untuk menyimpan dataset.

```
```{r}
#tm -> untuk membersihkan data
library(tm)
#vroom -> untuk meload dataset
library(vroom)
#here -> untuk menyimpan dataset
library(here)
```
```

Listing Program 3.1 Library Untuk Data Cleaning

3.2 Data Cleaning

Setelah semua library yang dibutuhkan untuk proses analisis data telah didefinisikan, maka proses selanjutnya yang akan dilakukan adalah proses pembersihan data. Pada proses ini, dilakukan pembersihan pada data seperti membersihkan URL, kemudian membersihkan new line, menghapus tanda baca seperti koma, titik, titik dua sehingga akan dihasilkan data yang bersih dari tanda baca maupun karakter lainnya. Selain pembersihan tanda baca, stopwords juga digunakan pada proses pembersihan data ini. Setelah data sudah berhasil dibersihkan, data akan disimpan ke dalam `clean_reviews.csv`

```
```{r load dataset}
dataset <- vroom(here('Amazon_Alexa_Review.csv'))

data_reviews <- dataset$verified_reviews
data_reviews2 <- Corpus(VectorSource(data_reviews))

=== Data Cleaning ===
Membersihkan URL
deleteURL <- function(x) gsub("http[^[:space:]]*", "", x)
cleanReview <- tm_map(data_reviews2, deleteURL)

#membersihkan New Line
deleteNewLine <- function(y) gsub("\n", "", y)
cleanReview <- tm_map(cleanReview, deleteNewLine)
```

```

Membersihkan tanda koma
replaceComma <- function(y) gsub(",", "", y)
cleanReview <- tm_map(cleanReview, replaceComma)

Membersihkan tanda titik dua
deleteColon <- function(y) gsub(":", "", y)
cleanReview <- tm_map(cleanReview, deleteColon)

Membersihkan tanda titik koma
deleteSemiColon <- function(y) gsub(";", " ", y)
cleanReview <- tm_map(cleanReview, deleteSemiColon)

Membersihkan tanda titik tiga
deleteEllipsis <- function(y) gsub("p...", "", y)
cleanReview <- tm_map(cleanReview, deleteEllipsis)

Membersihkan amp
deleteAmp <- function(y) gsub("&", "", y)
cleanReview <- tm_map(cleanReview, deleteAmp)

Membersihkan karakter
deleteUN <- function(z) gsub("@\\w+", "", z)
cleanReview <- tm_map(cleanReview, deleteUN)
remove.all <- function(xy) gsub("[^[:alpha:][:space:]]*", "", xy)

Membersihkan Tanda Baca
cleanReview <- tm_map(cleanReview, remove.all)
cleanReview <- tm_map(cleanReview, removePunctuation)
cleanReview <- tm_map(cleanReview, tolower)

Stopwords = readLines("stopwords.txt")
cleanReview <- tm_map(cleanReview, removeWords, Stopwords)
dataframe<-data.frame(text=unlist(sapply(cleanReview, `[`)),
stringsAsFactors=F)
View(dataframe)

Menyimpan data yang telah dibersihkan
write.csv(dataframe, file = 'clean_reviews.csv')

...

```

**Listing Program 3.2 Data Cleaning**

### 3.3 Barplot

Setelah dilakukan pembersihan data, maka proses selanjutnya adalah menampilkan data dalam bentuk barplot. Pada proses ini, diperlukan beberapa *library* yang mendukung proses ini diantaranya adalah library *e1071* yang digunakan untuk perhitungan naive bayes, kemudian library *caret* untuk melakukan klasifikasi data dan juga library *syuzhet* untuk `get_nrc`. Pada proses ini, data yang berada pada

clean\_reviews.csv ini akan diubah menjadi character(char) untuk dilakukan klasifikasi pada analisis sentimen.

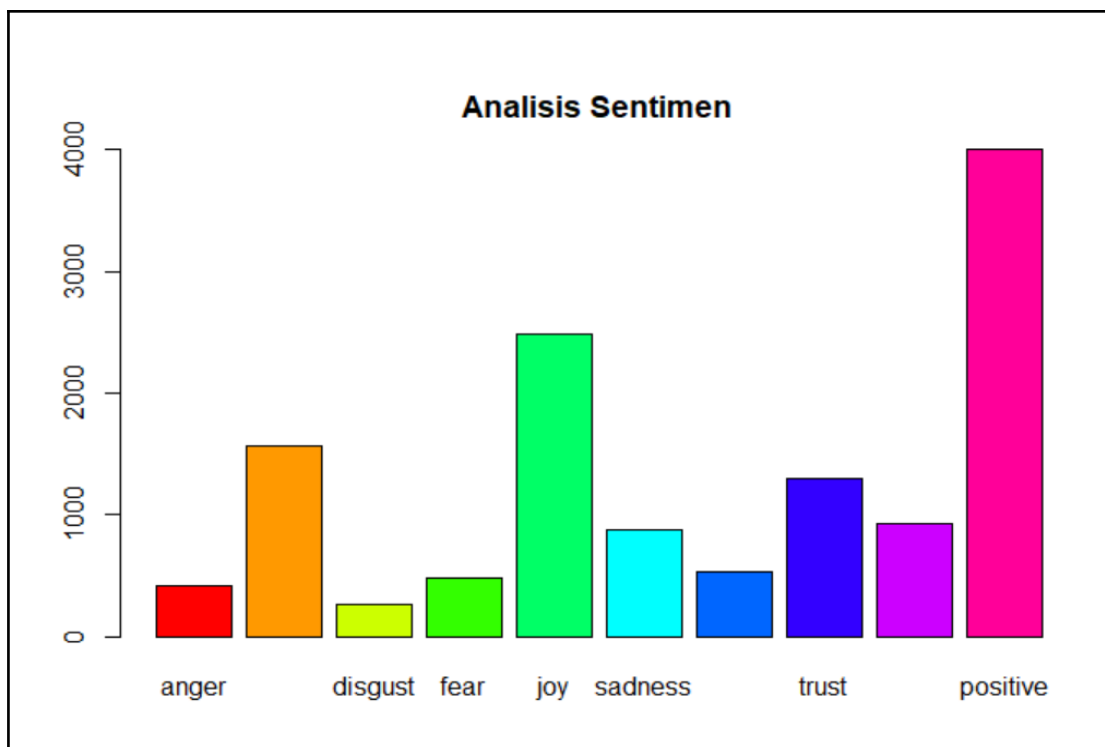
```
```{r Barplot}
# Library naive bayes
library(e1071)
# Library klasifikasi data
library(caret)
# Library untuk fungsi get_nrc
library(syuzhet)

dataUlasan <- read.csv("clean_reviews.csv", stringsAsFactors = FALSE)

# Mengubah text menjadi char
review <- as.character(dataUlasan$text)
emotion <- get_nrc_sentiment(review)

# Melakukan klasifikasi data
review_combine <- cbind(dataUlasan$text, emotion)
par(mar=rep(3,4))
a <- barplot(colSums(emotion), col=rainbow(10), ylab='count',
main='Analisis Sentimen')
brplt <- a
```
```

**Listing Program 3.3** Barplot



**Gambar 3.1** Keluaran Barplot

### 3.4 Word Cloud

Setelah menampilkan data dalam bentuk barplot, proses selanjutnya adalah penampilan wordcloudnya. Pada proses penampilan word cloud ini dibutuhkan juga beberapa library diantaranya adalah library tm, kemudian RTextTools, e1071, dplyr, caret dan yang terakhir adalah library wordcloud. Pada proses ini, data akan dibersihkan terlebih dahulu dengan menggunakan corpus untuk selanjutnya data dapat ditampilkan.

```
```{r WordCloud}
# Library untuk corpus dalam cleaning data
library(tm)
library(RTextTools)

# Library algoritma naivebayes
library(e1071)
library(dplyr)
library(caret)

#membaca file csv
df <- read.csv("clean_reviews.csv",stringsAsFactors = FALSE)
#melihat tipe dan juga struktur objek
glimpse(df)

#Mengatur seed generator bilangan acak R(berfungsi agar bilangan acak
R akan tetap sama walau diproduksi berulang kali)
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])

# Membersihkan kembali 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])

#Untuk klasifikasi data menggunakan data training dan data test
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]

#Kemudian teks akan diubah menjadi DTM(Document-Term Matrix) dengan
proses tokenization -> akan mengubah kalimat menjadi beberapa term(1
kata, 2 kata, dst).
dim(dtm.train)
fivefreq <- findFreqTerms(dtm.train,5)
length(fivefreq)
dtm.train.nb <- DocumentTermMatrix(corpus.clean.train,control =
list(dictionary=fivefreq))
#dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.clean.test,control =
list(dictionary=fivefreq))
dim(dtm.test.nb)
convert_count <- function(x){
  y <- ifelse(x>0,1,0)
  y <- factor(y,levels=c(0,1),labels=c("no","yes"))
  y
}

#Data train dan test naive bayes
trainNB <- apply(dtm.train.nb,2,convert_count)
testNB <- apply(dtm.test.nb,1,convert_count)

#menampilkan wordcloud
library(wordcloud)
wordcloud(corpus.clean, min.freq = 4, max.words=100, random.order=F,
colors=brewer.pal(8,"Dark2"))
```

```

**Listing Program 3.4 Word Cloud**



```

Alexa_review <- data.frame(Alexa_review)

#bagian UI
ui <- fluidPage(
 titlePanel("Analisis Sentimen Terhadap Amazon Alexa"),
 mainPanel(
 tabsetPanel(type = "tabs",
Data Review
tabPanel("Data Amazon Alexa", DT::dataTableOutput('tbl2')),
Data Clean
tabPanel("Data Clean", DT::dataTableOutput('tbl')),
Barplot
tabPanel("Barplot", plotOutput("scatterplot")),
Word Cloud
tabPanel("Wordcloud", plotOutput("Wordcloud"))
)
)
)

SERVER
server <- function(input, output) {

Output Data Tabel
output$tbl2 = DT::renderDataTable({
 DT::datatable(dataset <-
 vroom(here('Amazon_Alexa_Review.csv')), options =
 list(lengthChange = FALSE))
})

Output Data Clean
output$tbl = DT::renderDataTable({
 DT::datatable(Alexa_review, options = list(lengthChange =
 FALSE))
})

Output Plot
output$scatterplot <- renderPlot({
 par(mar=rep(3,4))
 a <-
barplot(colSums(emotion),col=rainbow(10),ylab='count',main='Sentimen
 Analisis'), height=400)
df <- read.csv("clean_reviews.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])

Membersihkan 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))
#dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.clean.test, control =
list(dictionary=fivefreq))
dim(dtm.test.nb)
convert_count <- function(x){
 y <- ifelse(x>0,1,0)
 y <- factor(y, levels=c(0,1), labels=c("no", "yes"))
 y
}
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 1, convert_count)

Output WordCloud
output$Wordcloud <- renderPlot({
 wordcloud(corpus.clean, min.freq =
 4, max.words=100, random.order=F, colors=brewer.pal(8, "Dark2"))
})
}

Program Shiny
shinyApp(ui = ui, server = server)
```

```

Listing Program 3.5 Aplikasi Shiny

Analisis Sentimen Terhadap Amazon Alexa

Data Amazon Alexa

Data Clean

Barplot

Wordcloud

Search:

	rating	date	variation	verified_reviews	feedback
1	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
2	5	31-Jul-18	Charcoal Fabric	Loved it!	1
3	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer a question correctly but Alexa says you got it wrong and answers the same as you. I like being able to turn lights on and off while away from home.	1
4	5	31-Jul-18	Charcoal Fabric	I have had a lot of fun with this thing. My 4 yr old learns about dinosaurs, i control the lights and play games like categories. Has nice sound when playing music as well.	1
5	5	31-Jul-18	Charcoal Fabric	Music	1
6	5	31-Jul-18	Heather Gray Fabric	I received the echo as a gift. I needed another Bluetooth or something to play music easily accessible, and found this smart speaker. Can't wait to see what else it can do.	1

Gambar 3.3 Data Amazon Alexa

Analisis Sentimen Terhadap Amazon Alexa

Data Amazon Alexa

Data Clean

Barplot

Wordcloud

Search:

...1

text

1	1	love echo
2	2	loved
3	3	sometimes playing game answer question correctly alexa wrong answers able lights home
4	4	lot fun yr learns dinosaurs control lights play games categories nice sound playing music
5	5	music
6	6	received echo gift bluetooth play music easily accessible found smart speaker cant wait else
7	7	cellphone features ipad alarm deaf hear alarm bedroom living reason herit fun random questions hear response smartbon politics
8	8	th ive purchased im getting house features offer specifily playing music echos controlling lights throughout house
9	9	looks
10	10	love ive listened songs havent heard childhood news weather information

Showing 1 to 10 of 3,150 entries

Previous

1

2

3

4

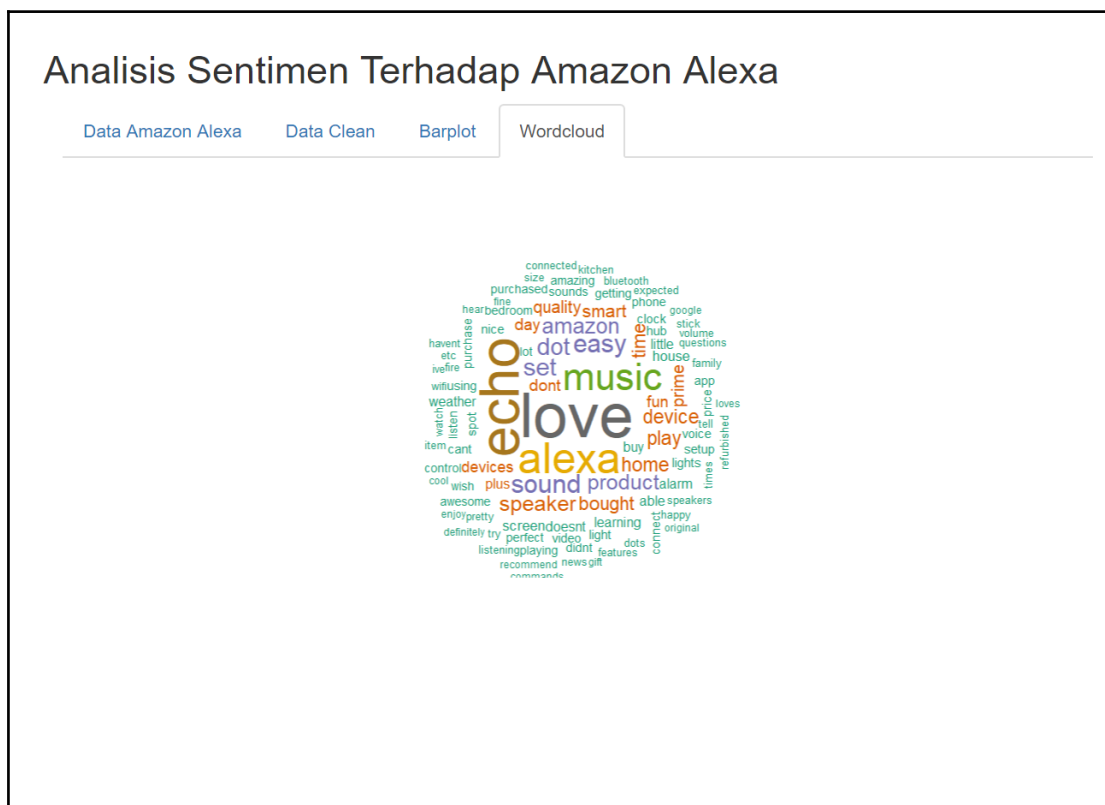
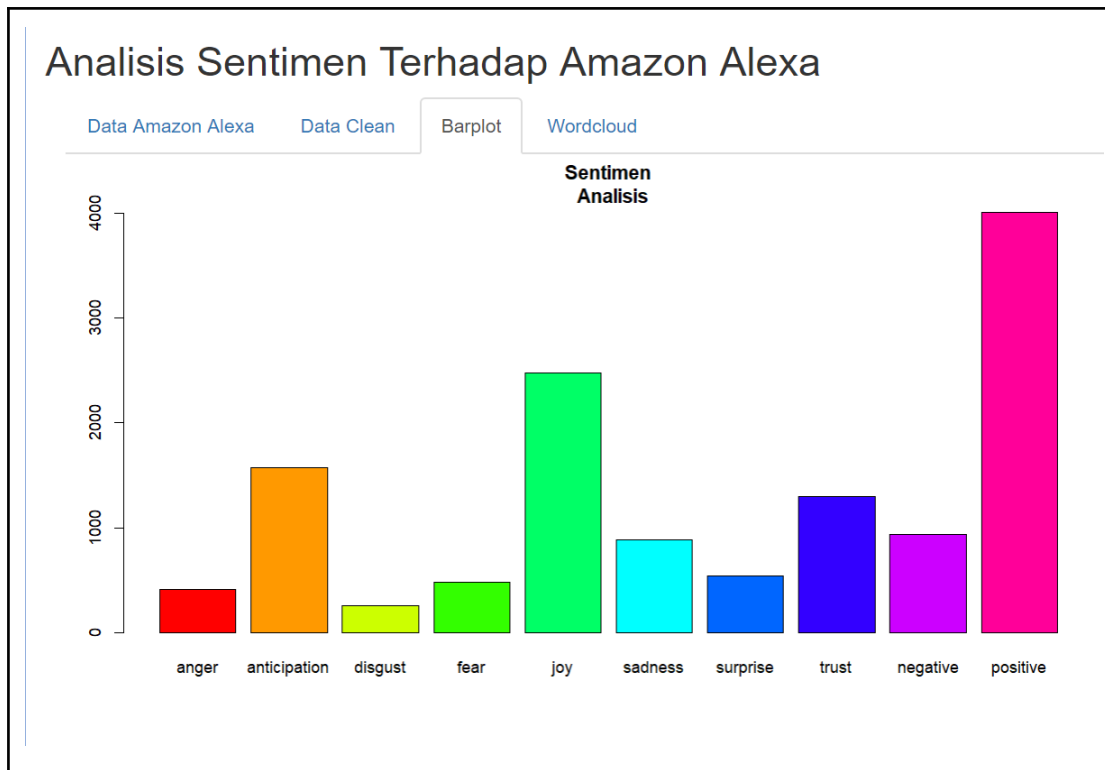
5

...

315

Next

Gambar 3.4 Data yang Telah Dibersihkan(*Clean Data*)



4. KESIMPULAN

Kesimpulan yang didapatkan dari hasil project analisis sentimen terhadap review pengguna alexa adalah sebagai berikut:

1. Program sentimen analisis ini dapat berjalan dengan baik, mulai dari pembersihan data, pengelompokan data, hingga hasil pada barplot dan word cloud.
2. Program menghasilkan analisis sentimen yang sesuai dengan yang diharapkan dimana dapat diketahui hasil yang bernilai positif maupun negatif.
3. Hasil pengujian pada program analisis sentimen ini menghasilkan kesimpulan juga bahwa review terbanyak berupa kata-kata positif dan kata-kata yang paling sering muncul pada review adalah love.