**LAPORAN PROYEK AKHIR**

**PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN REVIEW APLIKASI TRAVELOKA**

****

|  |  |
| --- | --- |
| MUHAMMAD RIZKI FIRMANSYAH | 123200064 |
| ARDHIAN KUSUMAYUDA | 123200144 |

**PROGRAM STUDI INFORMATIKA**

**JURUSAN INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2022**

**HALAMAN PENGESAHAN LAPORAN PROYEK AKHIR**



Disusun oleh :

Muhammad Rizki Firmansyah123200064

Ardhian Kusumayuda123200144

Telah Diperiksa dan Disetujui oleh Asisten Praktikum Data Science

Pada Tanggal : ................................

.

|  |  |  |
| --- | --- | --- |
| **Asisten Praktikum**  **Rivano Ardiyan Taufiq Kurniawan**  1232190084 |  | **Asisten Praktikum**  **Ega Erinovian Megananda Putra**  1232190113 |

# **DAFTAR ISI**

[**DAFTAR ISI** 3](#_Toc120711805)

[**1.** **PENDAHULUAN** 4](#_Toc120711806)

[**2.** **METODE DAN TEORI** 4](#_Toc120711807)

[**2.1 Teori** 5](#_Toc120711808)

[**2.1.1 Sentimen Analisis** 5](#_Toc120711809)

[**2.1.2 Naive Bayes** 5](#_Toc120711810)

[**2.2 Metode** 5](#_Toc120711811)

[**2.2.1 Data Collection** 5](#_Toc120711812)

[**2.2.2 Data Cleaning** 6](#_Toc120711813)

[**3.** **HASIL DAN PEMBAHASAN** 7](#_Toc120711814)

[**3.1 Data Collection** 7](#_Toc120711815)

[**3.2 Code Program** 7](#_Toc120711816)

[**3.3 Tampilan GUI** 14](#_Toc120711817)

[**4.** **KESIMPULAN** 16](#_Toc120711818)

[**5.** **PEMBAGIAN TUGAS** 17](#_Toc120711819)

# **PENDAHULUAN**

Seiring berkembangnya zaman, perkembangan teknologi khususnya teknologi informasi semakin maju dan canggih. Perkembangan ini mempermudah semua kalangan masyarakat di dunia, contoh hasil perkembangannya adalah internet. Internet saat ini telah hampir seluruh dunia dapat mengaksesnya, tidak hanya kalangan atas, namun seluruh lapis masyarakat dapat memakainya. Dengan kemudahan internet, masyarakat mampu mengakses berbagai fitur atau fasilitas online yang ada, salah satunya adalah aplikasi review tempat pariwisata yang bernama Traveloka.

Traveloka adalah perusahaan yang menyediakan layanan pemesanan tiket pesawat dan hotel secara daring dengan fokus perjalanan domestik di Indonesia. Didirikan pada tahun 2012 oleh Ferry Unardi, Derianto Kusuma, dan Albert Zhang. Traveloka yang merupakan salah satu perusahaan rintisan (startup) berstatus unicorn asal Indonesia ini mengembangkan layanannya pada pemesanan tiket kereta api, bus, penyewaan mobil, hingga aktivitas wisata. Sejak tahun 2015, Traveloka mulai berekspansi ke sejumlah negara di Asia Tenggara seperti Singapura, Malaysia, Vietnam, Thailand, dan Filipina. Traveloka juga merupakan pemain utama di kategori local services (saat ini terbatas pada wilayah tertentu), yang menyajikan pemesanan untuk beragam tempat rekreasi dan kegiatan, klinik kesehatan dan kecantikan, direktori kuliner serta layanan pesan antar makanan. Traveloka juga menawarkan solusi keuangan, pembayaran, dan asuransi untuk membantu masyarakat di Asia Tenggara memenuhi aspirasi gaya hidup mereka. Traveloka memiliki layanan customer service yang tersedia 24/7 dalam bahasa lokal, serta kemudahan bertransaksi dengan lebih dari 40 metode pembayaran. Aplikasi Traveloka telah diunduh lebih dari 100 juta kali, menjadikannya sebagai aplikasi pemesanan perjalanan dan gaya hidup paling populer di kawasan Asia Tenggara.

Walaupun Traveloka menjadi aplikasi pemesanan perjalanan dan gaya hidup paling populer di Asia Tenggara, terdapat berbagai review atau ulasan postifi maupun negatif serta netral tentang aplikasi ini, khususnya review yang ada di Google Play. Dengan adanya penelitian ini, diharapkan dapat memberikan gambaran dan membuat masyarakat lebih selektif serta berhati-hati ketika akan mengunduh / download dan menggunakan aplikasi Traveloka sebagai aplikasi pembantu ketika akan melakukan suatu perjalanan ke kota maupun negara lain.

# **METODE DAN TEORI**

Data Science adalah cabang ilmu yang mengintegrasikan elemen statistik, ilmu komputer, dan matematika untuk memecahkan masalah atau mengembangkan ide dengan menentukan algoritma yang akan memberikan hasil terbaik. Metode yang akan digunakan dalam proyek kami adalah metode Sentimen Analisis dengan Naive Bayes

## **2.1 Teori**

### **2.1.1 Sentimen Analisis**

Analisis sentimen adalah proses menganalisis teks digital untuk menentukan apakah nada emosional pesan tersebut positif, negatif, atau netral. Saat ini, perusahaan memiliki data teks dalam volume besar seperti email, transkrip obrolan dukungan pelanggan, komentar media sosial, dan ulasan. Alat analisis sentimen dapat memindai teks ini untuk secara otomatis menentukan sikap penulis terhadap suatu topik. Perusahaan menggunakan wawasan dari analisis sentimen untuk meningkatkan mutu layanan pelanggan dan meningkatkan reputasi merek.

Analisis sentimen memiliki manfaat antara lain :

* Memberikan wawasan yang objektif
* Membangun produk dan layanan yang lebih baik
* Menganalisis dalam skala besar
* Hasil waktu nyata

### **2.1.2 Naive Bayes**

Naive bayes merupakan metode pengklasifikasian paling populer digunakan dengan tingkat keakuratan yang baik. Banyak penelitian tentang pengklasifikasian yang telah dilakukan dengan menggunakan algoritma ini. Pada algoritma naive bayes pengklasifikasian tidak membutuhkan adanya pemodelan maupun uji statistik.

Naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Keuntungan algoritma naive bayes adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi naive bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar.

## **2.2 Metode**

### **2.2.1 Data Collection**

Data Collection atau pengumpulan data dilakukan dengan melakukan proses scrapping menggunakan Google Colab dengan bahasa Python. Data scrapping dilakukan dengan bersumber pada section aplikasi Traveloka di Google PlayStore dengan alamat aplikasi *com.traveloka.android*.

### **2.2.2 Data Cleaning**

Data yang telah dikumpulkan masih bersifat kotor dan diperlukan pembersihan data lebih lanjut agar hanya data-data yang penting yang diperlukan dalam proses selanjutnya. Maka dilakukan proses data cleaning sebagai berikut :

**2.2.2.1 Penghapusan Tanda Baca**

Data yang dikumpulkan terdapat banyak tanda baca yang harus dihapus agar data berisi kata-kata yang dibutuhkan saja tanpa tanda baca.

**2.2.2.2 Perbaikan Ejaan**

Pada proses pembersihan selanjutnya masih banyak ditemukan dari data yang tidak menggunakan ejaan yang disempurnakan (EYD), maka dalam hal ini perlu diperbaiki untuk mengurangi sifat ambiguitas.

**2.2.2.3 Penghapusan kata yang tidak dibutuhkan**

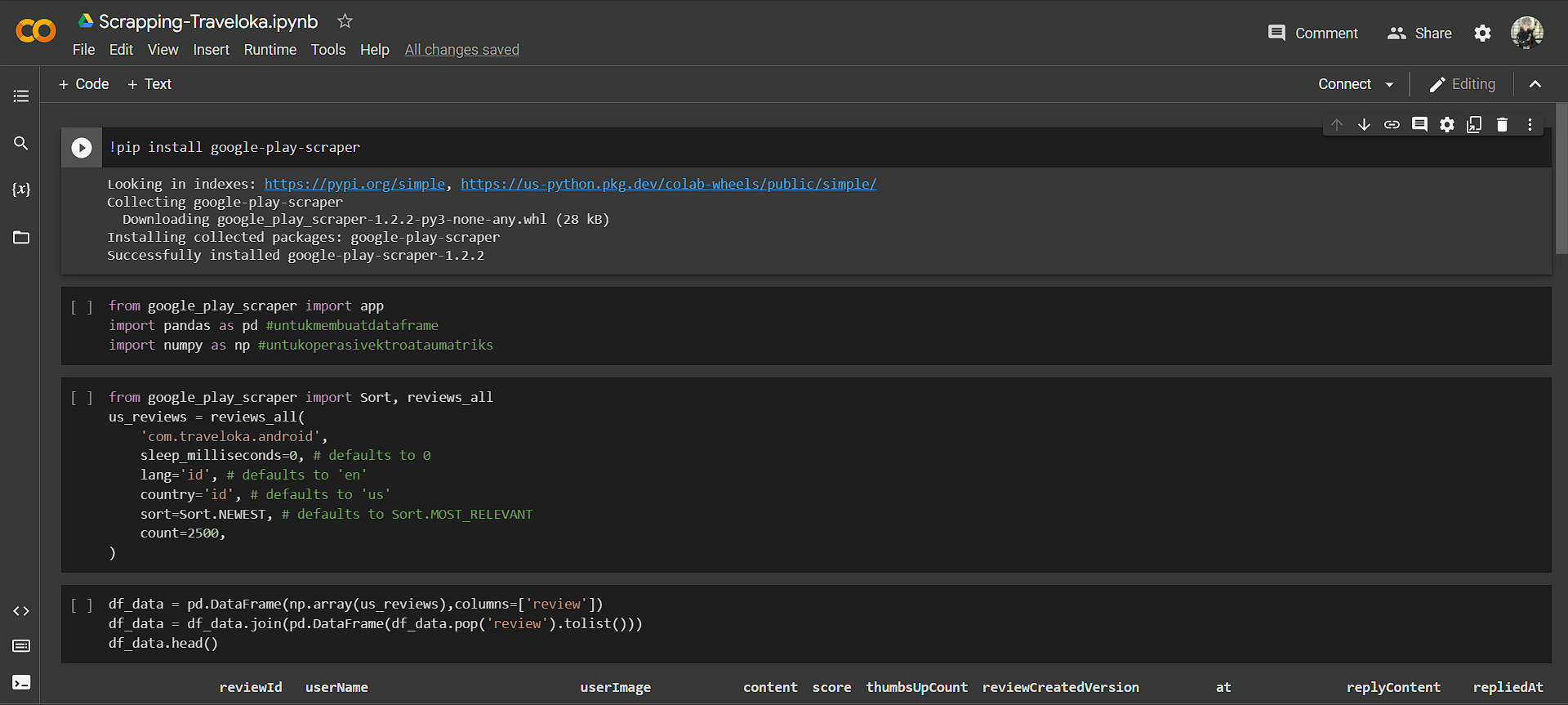
Pada data yang telah dikumpulkan terdapat kata - kata yang tidak dibutuhkan seperti kata "yang", "dan", "dari", "ini", "kita", "untuk", "nya", dll. Kata-kata ini dihapuskan dari data yang telah terkumpul.

**2.2.2.4 Pelabelan Data**

Data yang telah melewati proses sebelumnya diklasifikasikan menjadi kelas tertentu yaitu, kata positif, negatif, dan netral untuk dilakukan proses pemetaan suatu himpunan atribut objek pada label kelas tertentu.

# **HASIL DAN PEMBAHASAN**

## **3.1 Data Collection**

****

Gambar 3.1.1 Proses *scrapping* data

Pada gambar diatas terlihat proses *scrapping data* review aplikasi Traveloka menggunakan Google Colab dengan bahasa Python. Data discrapping dari alamat aplikasi Traveloka yaitu *com.traveloka.android* pada Google PlayStore. Setelah *scrapping*, didapatkan data sejumlah 163777 data. Namun dikarenakan jumlah yang terlalu banyak, maka kami hanya memilih data sebanyak 3000 data.

## **3.2 Code Program**

**Bagian 1.** Pertama-tama *Import* library yang akan digunakan.

|  |
| --- |
| ```{r}  library(tm)  library(wordcloud2)  library(vroom)  library(here)  library(RTextTools)  library(dplyr)  library(wordcloud)  library(shiny)  library(ggplot2)  library(plotly)  ``` |

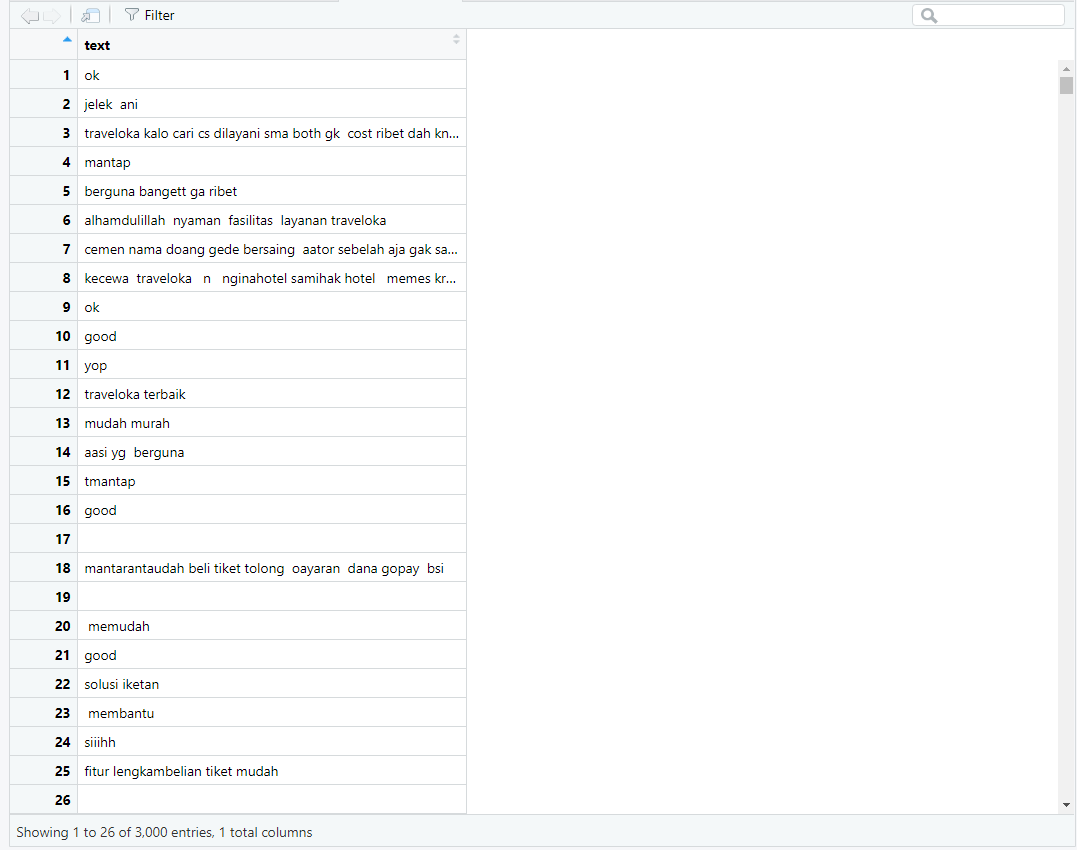
Listing 3.2.1 *Import* *library*

**Bagian 2.** Dilakukan langkah/proses *data cleaning* dengan membersihkan data data sehingga dapat dengan mudah dan lebih akurat ketika akan diolah. Data akan disimpan pada nama file “ReviewTraveloka-Clean.csv”.

|  |
| --- |
| ```{r Data Cleaning}  d <- vroom(here('ReviewTraveloka-3000.csv'))  ulasan <- d$content  ulasan1 <- Corpus(VectorSource(ulasan))  removeURL <- function(x) gsub("http[^[:space:]]\*", "", x)  reviewclean <- tm\_map(ulasan1, removeURL)  removeNL <- function(y) gsub("\n", " ", y)  reviewclean <- tm\_map(ulasan1, 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("&amp", "", 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)  MyStopWords <- readLines("stopwords-id.txt")  reviewclean <- tm\_map(reviewclean,removeWords,MyStopWords)  dataframe <- data.frame(text=unlist(sapply(reviewclean,`[`)),stringsAsFactors = F)  View(dataframe)  write.csv(dataframe,file = 'ReviewTraveloka-Clean.csv')  ``` |

Listing 3.2.2 Data Cleaning

Berikut adalah dataframe yang didapatkan setelah proses data cleaning



Gambar 3.2.1 *Dataframe* setelah proses *cleaning*

**Bagian 3.** Setelah mendapatkan data dicleaning proses selanjutnya adalah visualisasi data. Visualisasi data menggunakan bar plot atau diagram batang berdasarkan klasifikasi data dari fungsi analisis sentimen. Klasifikasi data memanfaatkan library caret dan fungsi get\_nrc dengan library syuzhet.

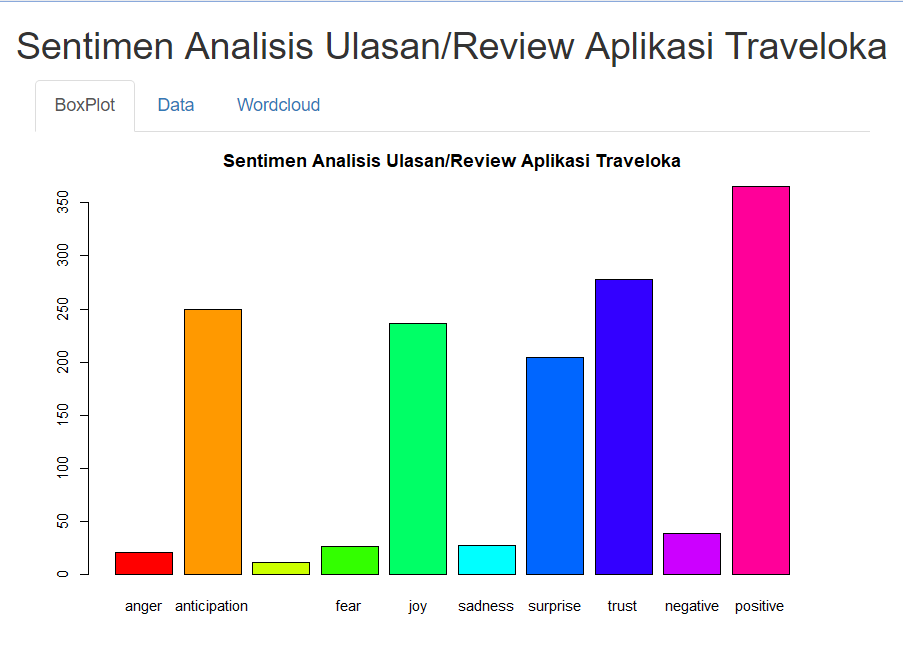
|  |
| --- |
| ```{r Sentimen Analisis Naive Bayes}  library(e1071) #Untuk Naive Bayes  library(caret) #untuk Klasifikasi Data  library(syuzhet) #untuk membaca fungsi get\_nrc  data\_traveloka <- read.csv("ReviewTraveloka-Clean.csv", stringsAsFactors = FALSE)  review <- as.character(data\_traveloka$text) #merubah text menjadi char  s <- get\_nrc\_sentiment(review)  review\_combine <- cbind(data\_traveloka$text,s) #klasifikasi Data  par(mar=rep(3,4))  a <- barplot(colSums(s), col=rainbow(10),ylab='count',main='Sentiment Analisis')  brplt <- a  ``` |

Listing 3.2.3 Visualisasi data *bar plot*

|  |
| --- |
| ```{r}  prosesdata <- read.csv("dataclean.csv", stringsAsFactors = FALSE)  review <- as.character(prosesdata$Reviews)  s <- get\_nrc\_sentiment(review, cl = NULL, language = "english", lowercase =TRUE)  # klasifikasi data  review\_combine <- cbind(prosesdata$Reviews, s)  par(mar = rep(3, 4))  datanew <- barplot(colSums(s), col = rainbow(10), xlab = 'emotion', ylab = 'count', main = 'Sentiment Analysis')  datanew  ``` |

Listing 3.2.4 Klasifikasi Data dan Sentimen Analisis

Hasil dari visualisasi data sebagai berikut:

  
Gambar 3.2.2 Visualisasi data berupa diagram batang

**Bagian 4.** Selanjutnya data divisualisasikan menggunakan wordcloud. Wordcloud menampilkan data kata-kata yang sering muncul atau yang sering digunakan pada review Aplikasi Traveloka.

|  |
| --- |
| ```{r}  require (corpus)  df<-read.csv("ReviewTraveloka-Clean.csv",stringsAsFactors = FALSE)  glimpse(df)  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, c("yang", "dan", "dari", "aasi", "ini", "kita", "untuk" ,"nya"))%>%  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)  wordcloud(corpus.clean,min.freq = 4,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))  kalimat2<-read.csv("ReviewTraveloka-Clean.csv",header=TRUE)  #skoring  kata.positif <- scan("kata-positif.txt",what="character",comment.char=";")  kata.negatif <- scan("kata-negatif.txt",what="character",comment.char=";")  score.sentiment = function(kalimat2, kata.positif, kata.negatif,  .progress='none')  {  require(plyr)  require(stringr)  scores = laply(kalimat2, function(kalimat, kata.positif,  kata.negatif) {  kalimat = gsub('[[:punct:]]', '', kalimat)  kalimat = gsub('[[:cntrl:]]', '', kalimat)  kalimat = gsub('\\d+', '', kalimat)  kalimat = tolower(kalimat)  list.kata = str\_split(kalimat, '\\s+')  kata2 = unlist(list.kata)  positif.matches = match(kata2, kata.positif)  negatif.matches = match(kata2, kata.negatif)  positif.matches = !is.na(positif.matches)  negatif.matches = !is.na(negatif.matches)  score = sum(positif.matches) - (sum(negatif.matches))  return(score)  }, kata.positif, kata.negatif, .progress=.progress )  scores.df = data.frame(score=scores, text=kalimat2)  return(scores.df)}  hasil = score.sentiment(kalimat2$text, kata.positif, kata.negatif)  #mengubah nilai score menjadi sentimen  hasil$klasifikasi<- ifelse(hasil$score<0, "Negatif",ifelse(hasil$score==0,"Netral","Positif"))  hasil$klasifikasi  #menukar urutan baris  data <- hasil[c(3,1,2)]  #View(data)  write.csv(data, file = "datalabel.csv")  ```  ```{r}  library(dplyr)  class\_sentiment <- data.frame(negative=s$negative,positive=s$positive)  klasifikasi <- mutate(class\_sentiment, text\_sentiment = ifelse((class\_sentiment$negative != class\_sentiment$positive),  ifelse(class\_sentiment$negative!=0,print("negative"),  print("positive")),print("neutral")))  data\_traveloka <- data.frame(text=review,s=klasifikasi$text\_sentiment)  View(data\_traveloka)  ``` |

Listing 3.2.5 Visualisasi Data dengan Worldcloud

Hasil visualisasi data berupa wordcloud:



Gambar 3.2.2 Wordcloud berdasarkan olahan dataset

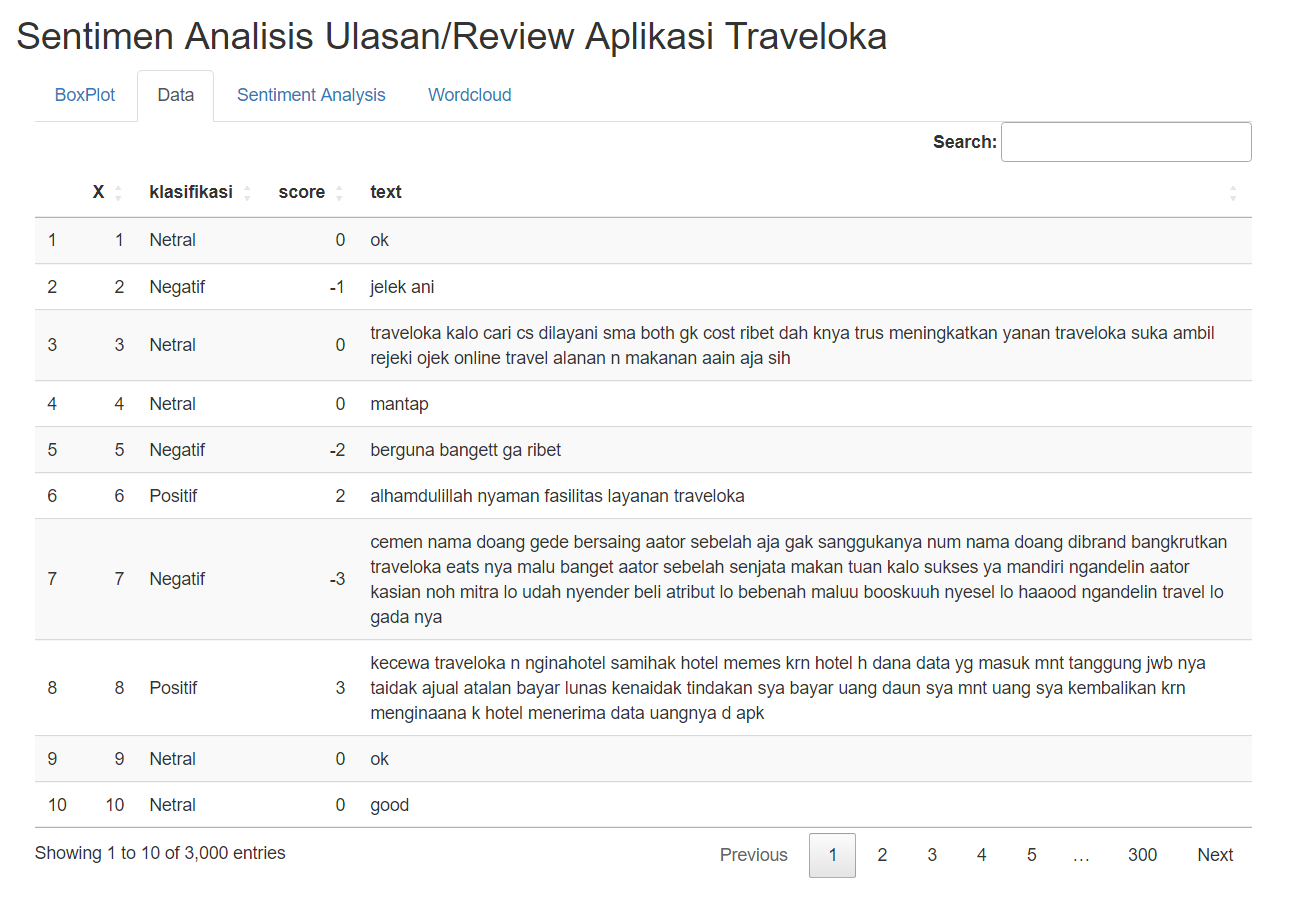
**Bagian 5.** Kemudian dibuat suatu GUI atau user inferface berdasarkan dataset yang telah diolah dan dilakukan visualisasi data di GUI tersebut,

|  |
| --- |
| ```{r}  library(syuzhet) #untuk membaca fungsi get\_nrc  dataLabel<- read.csv("datalabel.csv")  ui <- fluidPage(  titlePanel("Sentimen Analisis Ulasan/Review Aplikasi Traveloka"),  mainPanel(    tabsetPanel(type = "tabs",  tabPanel("BoxPlot", plotOutput("scatterplot")),  # Plot  tabPanel("Data", DT::dataTableOutput('tbl1')),  # Output Data Dalam Tabel  tabPanel("Wordcloud", plotOutput("Wordcloud"))  )  )  )  # SERVER  server <- function(input, output) {    # Output Data  output$tbl1 = DT::renderDataTable({  DT::datatable(dataLabel, options = list(lengthChange = FALSE))  })    output$scatterplot <- renderPlot({produk\_dataset<-read.csv("ReviewTraveloka-Clean.csv",stringsAsFactors = FALSE)  review <-as.character(produk\_dataset$text)  s<-get\_nrc\_sentiment(review)  review\_combine<-cbind(produk\_dataset$text,s)  par(mar=rep(3,4))  barplot(colSums(s),col=rainbow(10),ylab='count',main='Sentimen Analisis Ulasan/Review Aplikasi Traveloka')  }, height=400)  output$Wordcloud <- renderPlot({  set.seed(20)  df<-df[sample(nrow(df)),]  df<-df[sample(nrow(df)),]  glimpse(df)  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,]  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)  wordcloud(corpus.clean,min.freq = 4,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))  })  }  shinyApp(ui = ui, server = server)  ``` |

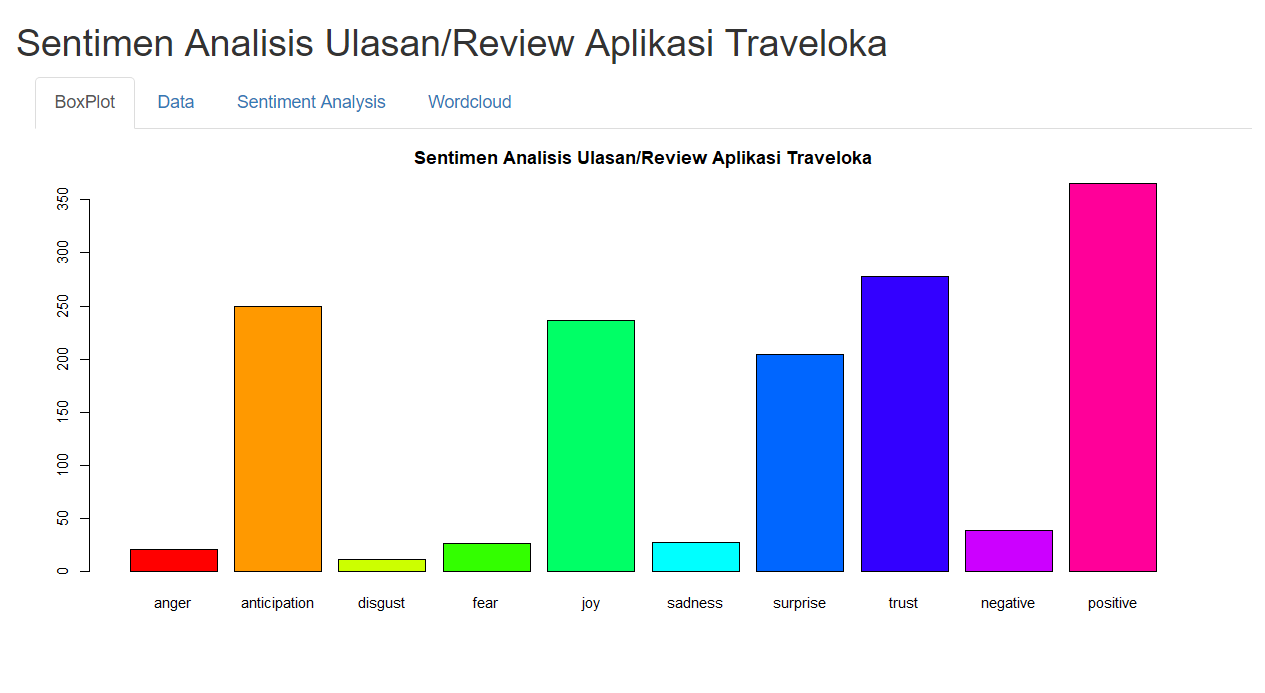
Listing 3.2.6 Visualisasi Data yang diolah dengan GUI Shiny

## **3.3 Tampilan GUI**

Library shiny digunakan untuk membuat tampilan GUI, berikut tampilannya:

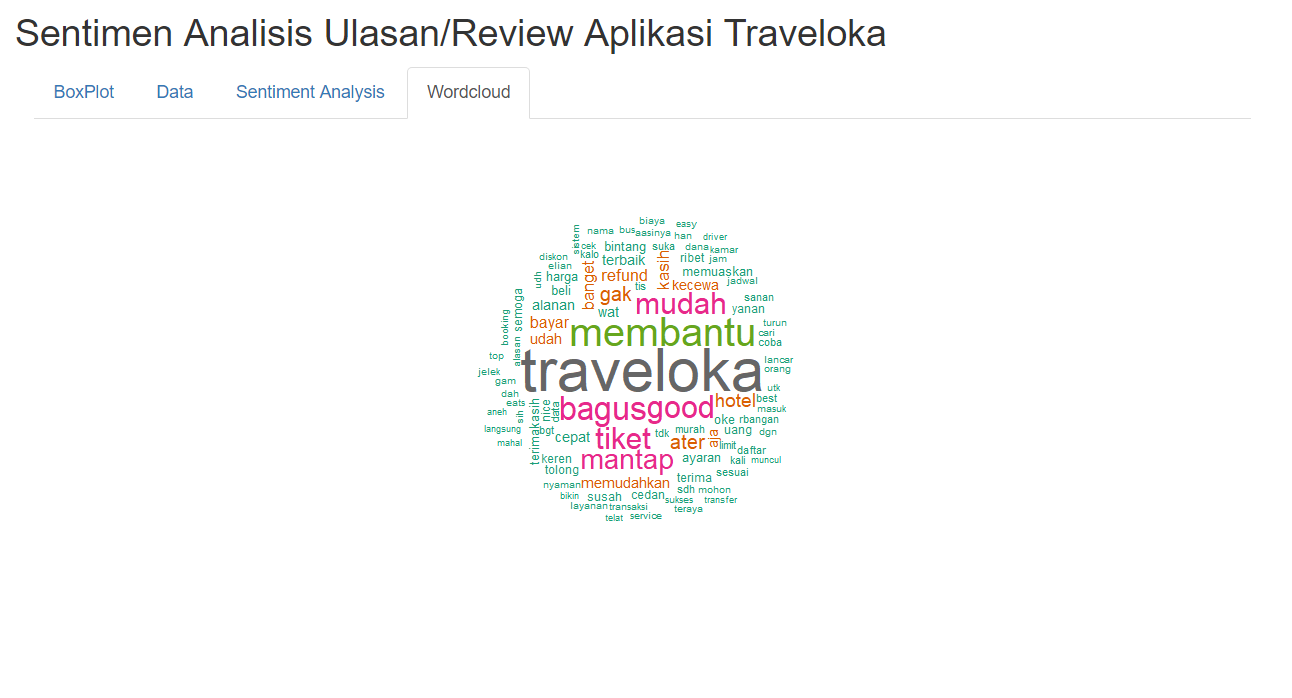
****

Gambar 3.3.1 Tampilan data

****Gambar 3.3.2 Tampilan Diagram batang/*bar plot*

****

Gambar 3.2.3 Tampilan Sentimen Analisis

****

Gambar 3.2.4 Tampilan Wordcloud

# **KESIMPULAN**

Berdasarkan data yang telah diolah didapatkan pernyataan bahwa terdapat lebih banyak komentar positif daripada komentar negatif atas review aplikasi Traveloka sehingga pengguna/user dapat dikatakan puas atas adanya aplikasi Traveloka dengan segala manfaat dan fitur didalamnya.

Berdasarkan pernyataan diatas, dapat disimpulkan bahwa aplikasi Traveloka merupakan aplikasi yang bermanfaat karena aplikasi itu sendiri dan fitur didalamnya sehingga memiliki banyak review positif bagi penggunanya. Namun, disisi lain masih terdapat berbagai ulasan negatif dan netral dimana hal itu tentu menjadi suatu hal yang harus diperhatikan oleh pihak Traveloka dalam mengelola aplikasi Traveloka kedepannya.

# **PEMBAGIAN TUGAS**

|  |  |  |
| --- | --- | --- |
| **No.** | **Aktivitas** | **Pelaksana** |
| 1. | Penentuan tema dan judul proyek | Rizki , Ardhian |
| 2. | Perancangan konsep dasar dan konsep utama proyek | Rizki , Ardhian |
| 3. | Pencarian dataset dan scrapping | Ardhian |
| 4. | Pembuatan program | Rizki , Ardhian |
| 5. | Pembuatan laporan | Rizki, Ardhian |
| 6. | Penyempurnaan laporan | Rizki |