**LAPORAN PROYEK DATA SCIENCE**

**ANALISIS SENTIMEN REVIEW LAZADA**

****

**DISUSUN OLEH:**

**Maulana Arya Wisnu Wardhana 123200138**

**Farras Alam Majid 123200044**

**PROGRAM STUDI INFORMATIKA JURUSAN INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2022**

1. **Pendahuluan**

Kemajuan teknologi adalah sesuatu yang tidak bisa dihindari dalam kehidupan ini, karena kemajuan teknologi akan berjalan sesuai dengan kemajuan ilmu pengetahuan. Setiap inovasi diciptakan untuk memberikan manfaat positif bagi kehidupan manusia.

Teknologi juga memberikan banyak kemudahan, serta sebagai cara baru dalam melakukan aktivitas manusia. Manusia juga sudah menikmati banyak manfaat yang dibawa oleh inovasi-inovasi teknologi yang telah dihasilkan dalam dekade terakhir ini.

Kemajuan Teknologi informasi dan Komunikasi telah membuka tahap baru bagi masyarakat untuk memperoleh informasi secara otonon. Perkembangan teknologi informasi dan komunikasi merupakan suatu yang harus ada dan diikuti oleh masyarakat modern saat ini.

Perkembangan teknologi dan komunikasi yang begitu pesat, telah memunculkan adanya sebuah media baru. Keberadaan media baru ini diantaranya adalah munculnya internet. Internet sebagai sebuah produk teknologi komunikasi, meski sudah berkembang sejak puluhan tahun yang lalu, namun hingga saat ini keberadaannya semakin dibutuhkan oleh hampir semua masyarakat dunia.

Salah satu sarana teknologi yang semakin berkembang saat ini adalah media belanja masyarakat, yang mana dahulu masyarakat harus pergi ke toko atau pasar untuk melakukan transaksi

Sekarang dengan adanya e-commerse masyarakat lebih mudah belanja dari mana pun, dan kapanpun itu.Hal itu dipermudah dengan semakin majunya teknologi yang dapat memudahkan masyarakat untuk mendapatkan barang.

1. **Metode**

Data science adalah sebuah ilmu yang menggabungkan aspek statistika, ilmu komputer, dan matematika dalam menyelesaikan permasalahan atau suatu gagasan 2 dengan menentukan algoritma agar hasil yang diperoleh optimal. Dalam proyek kami akan menggunakan Metode Naïve Bayes

* 1. **Business Understanding**

Business Understanding Dalam tahapan ini yang diharus dilakukan, yaitu menentukan permasalahan yang akan diselesaikan. Dalam hal ini kami mendapatkan suatu permasalahan, yaitu:

* Banyak masyarakat yang saat ini lebih sering menggunakan internet dalam kehidupan sehari – hari.
* Dengan adanya hal tersebut, masyarakat lebih mudah untuk mencari barang yang diinginkan diinternet
  1. **Data Understanding**

Dataset yang digunakan bersumber dari kaggle.com. Link dataset: <https://www.kaggle.com/datasets/grikomsn/lazada-indonesian-reviews>

* 1. **Data Preparation**

Pada tahap ini dilakukan proses untuk membangun dataset yang akan digunakan dalam tahap pemodelan, termasuk proses seleksi data, membersihkan data, dan menggabungkan data.

* 1. **Modelling**

Modeling dari project ini adalah tentang bagaimana review Lazada apakah memiliki respon positif atau negative.

Untuk mengetahui pertanyaan tersebut maka perlu dilakukan yaitu membuat model data yang sudah diperbaiki dengan menggunakan algoritma metode Naïve Bayes.

* 1. **Evaluation**

Dari evaluasi proyek yang telah dilakukan , dapat diketahui apakah model yang digunakan sudah dapat mengatasi permasalahan yang telah dilakukan tersebut.

* 1. **Deployment**

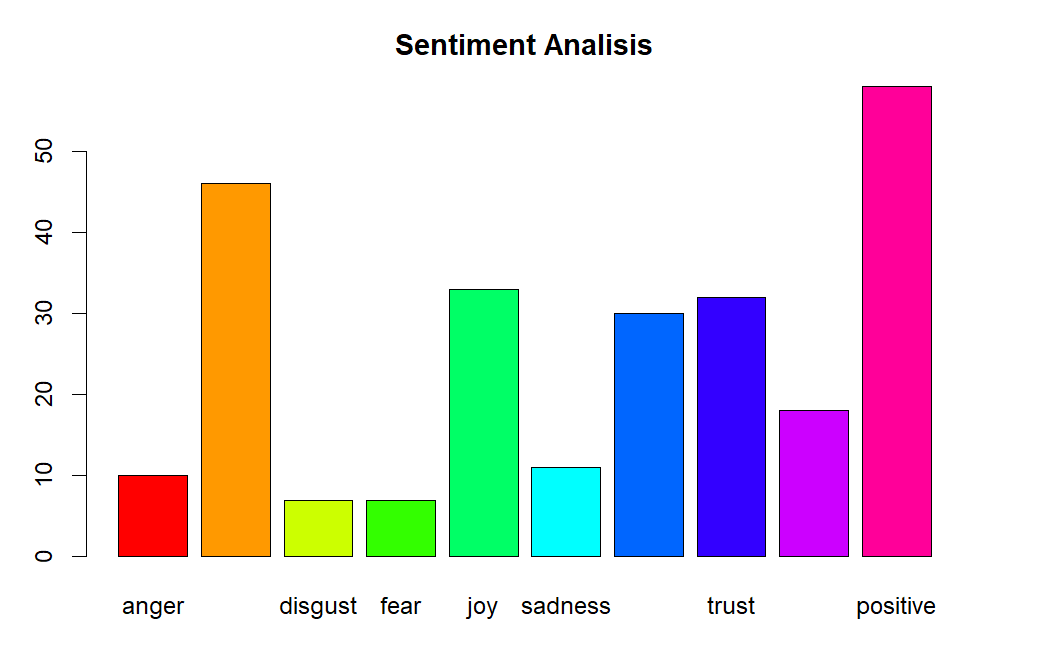
Saat output dari modeling yang sudah dikembangkan dapat diketahui apakah model tersebut sudah siap diterapkan atau digunakan di lingkungan teknologi ataupun bisnis.

1. **HASIL DAN PEMBAHASAN**

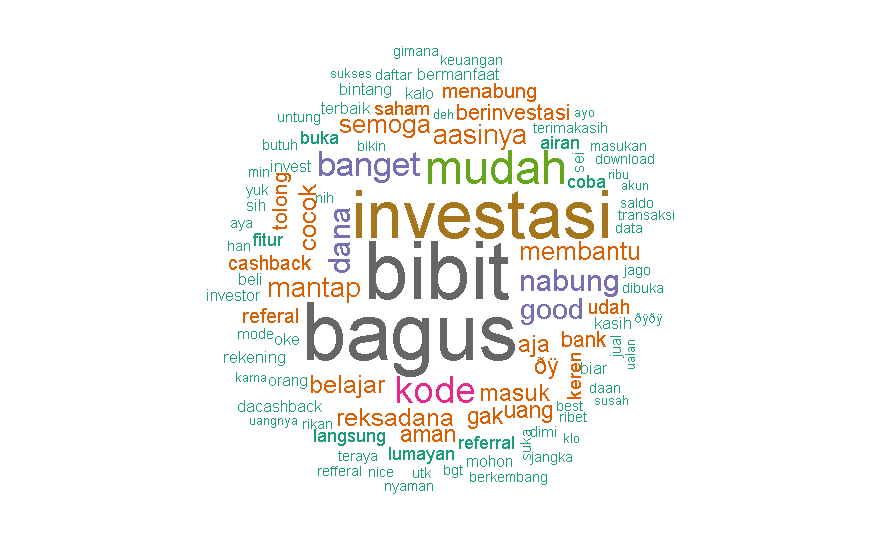
**3.1 Code Program**

Import library yang akan digunakan terlebih dahulu, kemudian set data frame diambil dari dataset yang digunakan dalam kasus ini.

|  |
| --- |
| library(wordcloud2)  library(vroom)  library(here)  library(RTextTools)  library(dplyr)  library(wordcloud)  library(shiny)  library(ggplot2)  library(plotly)  ```  ```{r Sentimen Analisis Naive Bayes Classifier}  library(e1071) #Untuk Naive Bayes  library(syuzhet) #untuk membaca fungsi get\_nrc  data <- read.csv("ReviewClean.csv", stringsAsFactors = FALSE)  review <- as.character(data$text) #merubah text menjadi char  s <- get\_nrc\_sentiment(review)  review\_combine <- cbind(data$text,s) #klasifikasi Data  par(mar=rep(3,4))  a <- barplot(colSums(s), col=rainbow(10),ylab='count',main='Sentiment Analisis')  brplt <- a  ```  Memanggil library tambahan yang akan digunakan untuk penggunaan corpus dalam proses cleaning data selanjutnya, Mengatur seed generator bilangan acak R, yang berguna untuk membuat simulasi atau objek acak yang dapat direproduksi.  ```{r}  require (corpus)  df<-read.csv("ReviewClean.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","lazada","null", "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))  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("ReviewClean.csv",header=TRUE)  #skoring  kata.positif <- scan("positip.txt",what="character",comment.char=";")  kata.negatif <- scan("negatip.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")  ```  Pengolahan serta pemodelan dari sebuah data yang telah di olah sebelumnya hingga menampilkan pada GUI dengann Library Shiny  ```{r}  library(syuzhet) #untuk membaca fungsi get\_nrc  dataLabel<- read.csv("datalabel.csv")  ui <- fluidPage(  titlePanel("Sentimen Analisis LAZADA"),  mainPanel(    tabsetPanel(type = "tabs",  tabPanel("Plot", plotOutput("scatterplot")),  # Plot  tabPanel("Data Bersih", 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("ReviewClean.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=c('gray'),ylab='count',main='Sentimen Analisis Ulasan Bibit')  }, 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)  ``` |

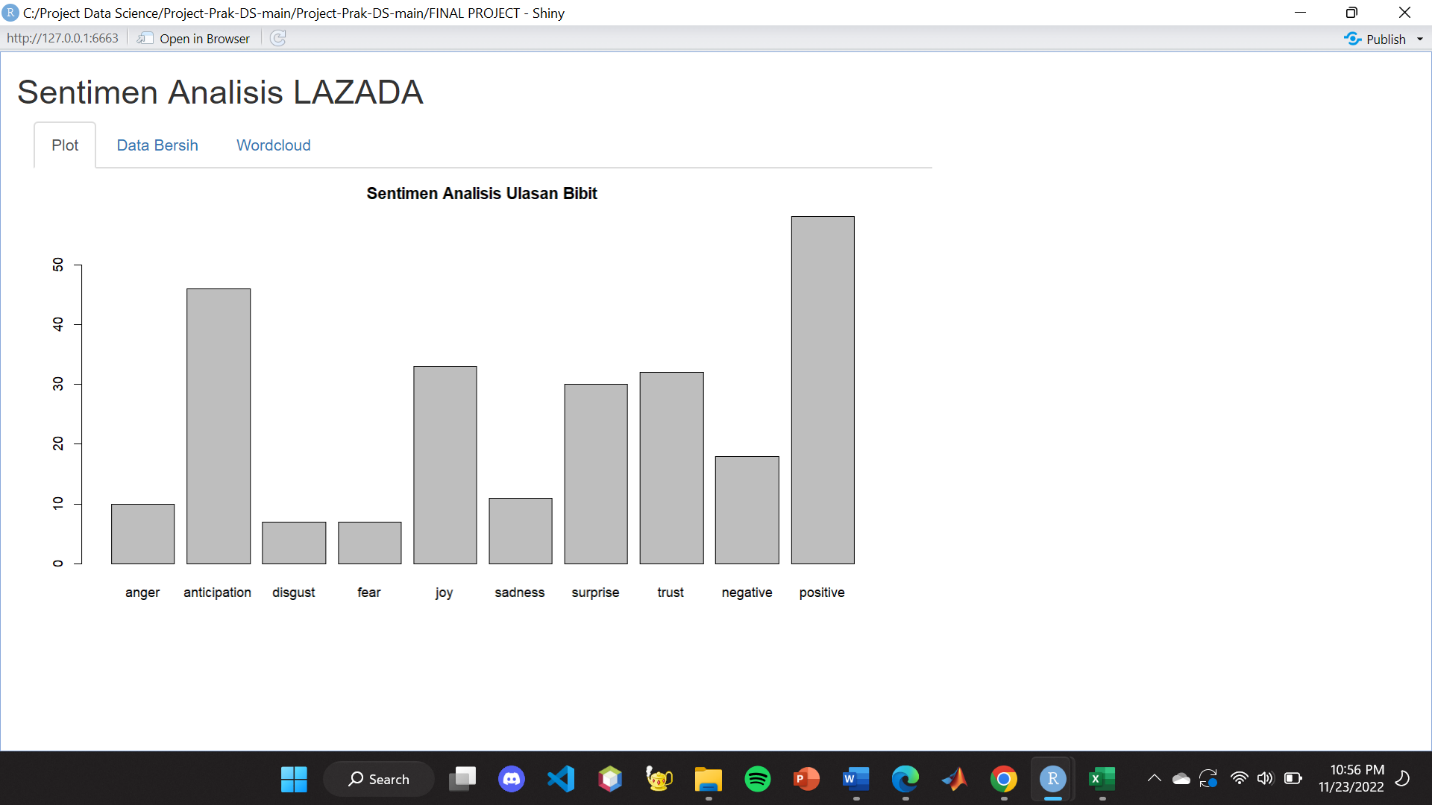


**Visualisasi data berupa diagram batang**

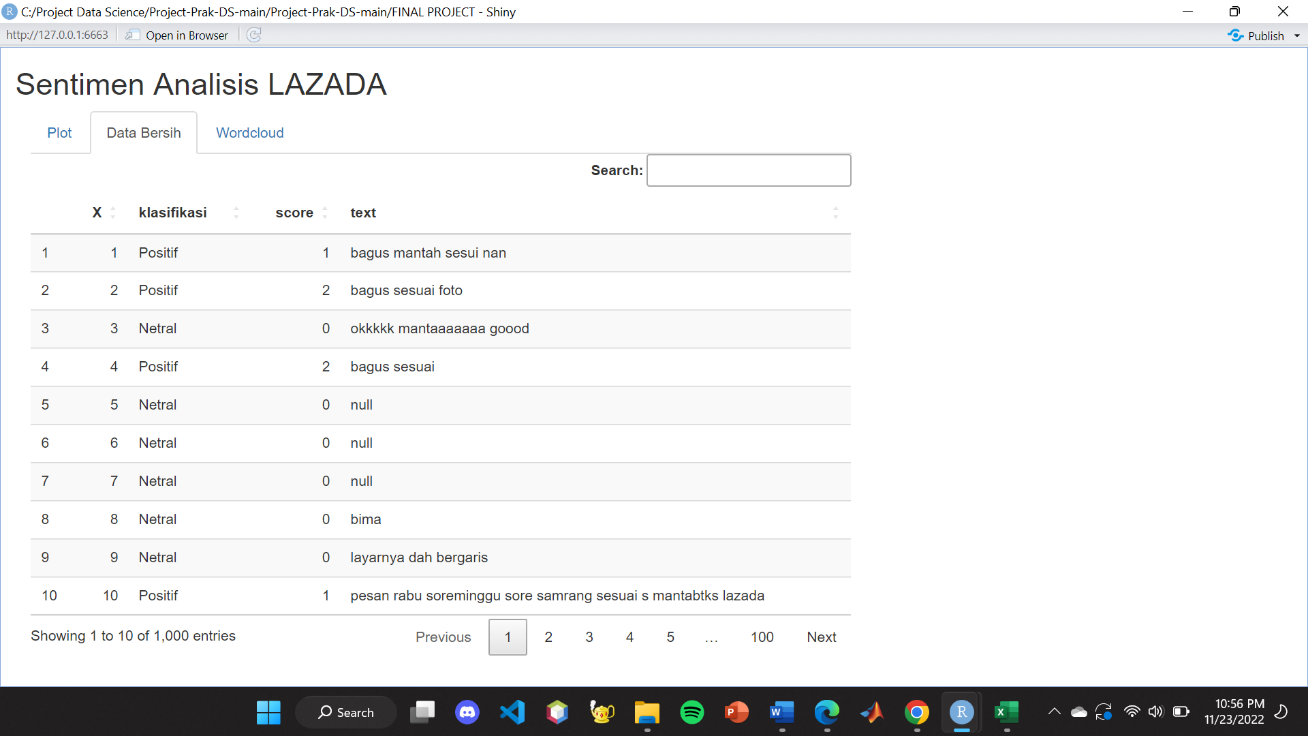
****

**Wordcloud berdasarkan olahan dataset**

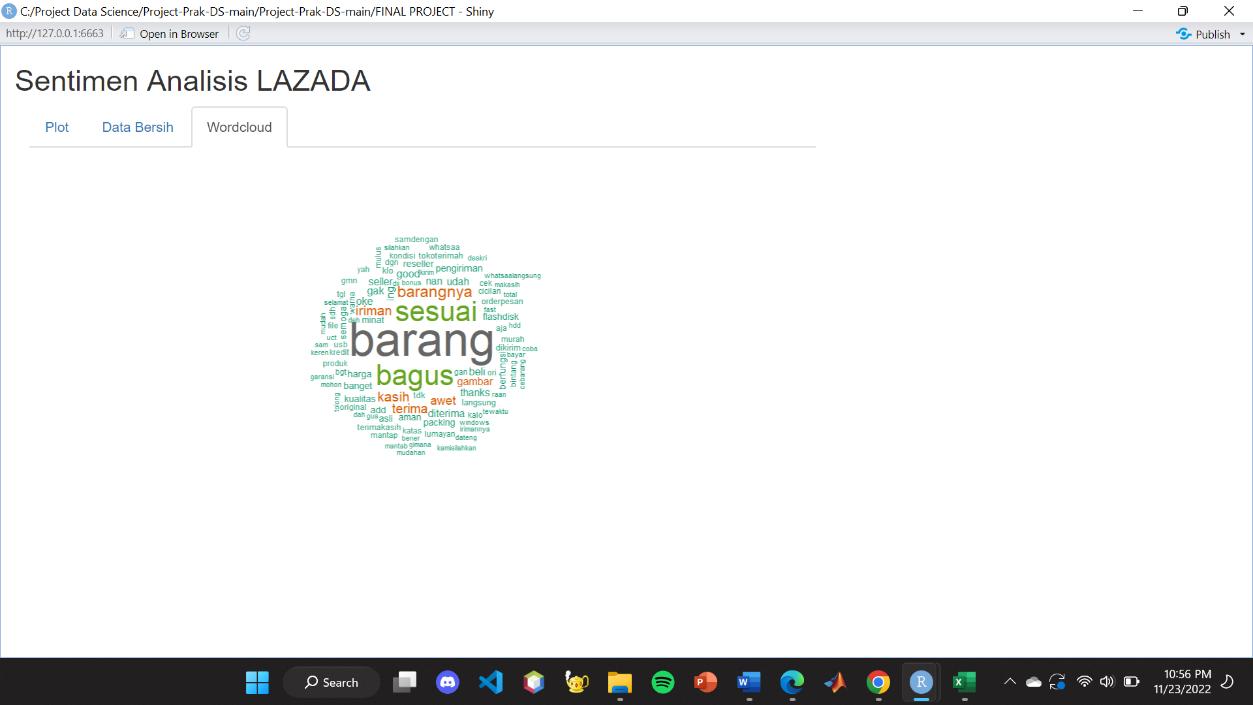
**Tampilan GUI**

****Library shiny digunakan untuk membuat tampilan GUI, berikut tampilannya:

**Tampilan Plot**

****

**Tampilan Data Bersih**

****

**Tampilan Wordcloud**

1. **KESIMPULAN**

Dari hasil berdasarkan data yang telah diolah sehingga mendapatkan pengklasifikasian:

* Dari ulasan yang telah didapat, bahwa terdapat banyak komentar positif dari pada ulasan/ komentar lainnya.

Berdasarkan hal tersebut dapat disimpulkan bahwa pengguna aplikasi Lazada memiliki nilai yang positif dimana dengan penggunaan aplikasi ini terdapat lebih banyak manfaat daripada keburukannya.

1. **PEMBAGIAN TUGAS**

|  |  |  |
| --- | --- | --- |
| No. | Aktivitas | Penanggung Jawab |
| 1 | Judul | Arya |
| 2 | Konsep Proyek | Farras |
| 3 | Cari Dataset | Arya |
| 4 | Pembuatan Program | Arya, Farras |
| 5 | Laporan | Farras |