

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE**

Analisis Sentimen Kepopuleran Cryptocurrency



Ahmad Bagus Nur Alfandi 123190120
Abigail Rakha Bhagaskara 123190144

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

1. PENDAHULUAN

1.1 Latar Belakang

Di era globalisasi dan masa pandemi saat ini, ilmu pengetahuan dan teknologi berkembang dengan sangat pesat, khususnya di bidang informasi dan komunikasi. Salah satunya adalah kemajuan di bidang komputer. Perkembangan teknologi ini memberikan banyak manfaat dalam kehidupan manusia karena dapat mempermudah pekerjaan sehari-hari sehingga lebih efektif dan efisien. Peranan teknologi informasi ini berpengaruh terhadap perkembangan daya saing suatu organisasi untuk memberikan layanan yang terbaik. Berdasarkan hal tersebut di berbagai instansi dan perusahaan terus diupayakan untuk meningkatkan sumber daya manusia secara sistematis dan terarah agar mampu mengimbangi sekaligus menguasai sains dan teknologi pada masa kini dan masa yang akan datang.

Di Indonesia perkembangan IT telah membawa Indonesia ke arah revolusi industri 4.0 yang diawali dengan revolusi internet. Seiring dengan perkembangan ini, dunia IT sedang diramaikan oleh machine learning dan Artificial Intelligence (AI). Adanya peningkatan ini membuat data mining dan analisa terhadap data menjadi fokus utama perusahaan untuk memberikan konteks terhadap arah perusahaan. Artinya, aktivitas ini memerlukan ilmu pengolahan data yang sering disebut dengan Data Science. Data science atau ilmu data merupakan ilmu terapan baru yang perkembangannya dituntut oleh meningkatnya penggunaan teknologi secara signifikan. Data science digunakan oleh perusahaan maupun instansi tertentu untuk melakukan analisis data yang tidak bisa dilakukan dengan metode sederhana.

Maka dari itu, disini kami menciptakan program Analisis Sentimen Kepopuleran Cryptocurrency. Kami memilih tema tersebut karena Cryptocurrency saat ini sedang hangat-hangat nya menjadi perbincangan dunia. Sehingga, diharapkan program kami dapat menganalisis respon dan tanggapan orang-orang terhadap Cryptocurrency.

1.2 Tujuan

Tujuan dari proyek Analisis Sentimen Kepopuleran Cryptocurrency ini adalah untuk menilai ekspresi emosi dari masyarakat seperti positif, netral, dan negatif serta menyelesaikan tugas akhir dari mata kuliah Praktikum Data Science.

2. METODE

2.1 Pengambilan data

Proses pengambilan data melalui api dari twitter sehingga setiap kali melakukan proses analisis akan terdapat perbedaan karena data yang diambil *up to date*.

2.2 Preprocessing

Setelah mendapatkan data kemudian dilakukan replacement terhadap tanda (<, >, /, :, ;, http\\w+, *) atau emoticon yang tidak berguna dengan spasi dan melanjutkan dengan tahap preprocessing. Tujuan dari tahap ini adalah untuk membersihkan data agar tidak menyulitkan proses analisis sentimen nantinya.

Yang dibersihkan yaitu :

- a) Remove Punctuation, berfungsi untuk menghilangkan tanda baca yang masih tersisa saat replacement.
- b) Remove Numbers, berfungsi untuk menghilangkan angka.
- c) Remove stripWhiteSpace, berfungsi untuk menghilangkan ekstra spasi dari replacement.
- d) Remove stopWords, berfungsi untuk menghilangkan kata umum yang tidak memiliki makna.
- e) Content Transformer, berfungsi untuk mengubah semua huruf menjadi huruf kecil.

2.3 Naïve Bayes

Data frame yang telah dibersihkan akan dibaca dan ditambahkan sebuah variable baru. Variable tersebut berfungsi untuk menampung nilai yang telah diberikan pada data review. Bila nilai data review kurang dari 0 maka review tersebut merupakan review negatif dan begitu juga sebaliknya. Tak lupa memberikan bobot pada tiap-tiap kata. Pembobotan dilakukan untuk mendapatkan nilai dari data review yang berhasil dibersihkan. Setelah membuat model Naïve Bayes untuk memprediksi data frame yang telah diberi nilai, yang terakhir menyimpan hasil prediksi ke dalam data frame.

2.4 Shiny

Shiny merupakan sebuah paket pada R Studio untuk membuat website. Shiny menggabungkan antara komputasi statistika R dan interaksinya dengan website modern. Shiny sendiri terdiri dari 4 bagian yaitu :

a) Global

berfungsi untuk mencantumkan library-library yang akan digunakan. Selain mencantumkan library, bagian ini juga bisa digunakan untuk memuat data yang akan digunakan, menentukan source file, proses autentikasi, dan semua pengaturan yang bersifat global untuk kepentingan penelitian ini.

b) User Interface

Digunakan untuk mendefinisikan tampilan web dari penelitian ini dan memuat seluruh fungsi input dan output. Pada bagian ini digunakan dashboardpage untuk halamannya yang berisi dashboardheader untuk bagian atas dari halaman (judul), dashboardsidebar untuk menu bagian kiri, dan dashboardbody untuk isi website (analisis sentimen, wordcloud)

c) Server

Merupakan fungsi yang mendefinisikan logika analysis dari sisi server pada penelitian ini.

d) ShinyApp

Merupakan fungsi untuk memanggil UI dan server yang telah dibuat untuk dijalankan.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada Analisis Sentimen terhadap Kepopuleran Cryptocurrency.

Pada listing program 3.1 yaitu melakukan inisialisasi terhadap library yang diperlukan.

```
```{r}
library(tm)
library(wordcloud2)
library(twitterR)
library(rtweet)
library(shiny) #package shiny
library(syuzhet) #package analisis sentimen
library(wordcloud) #package wordcloud
library(vroom)
library(here)
library(dplyr)
library(ggplot2)
library(RColorBrewer)
library(plyr)
library(RTextTools)
```

**Listing Program 3.1** Inisialisasi Library yang diperlukan

Kemudian kami menggunakan api key dari twitter untuk mengambil data dengan key search Cryptocurrency dan dengan data maksimal adalah 1000. Kemudian data yang sudah diminning akan disimpan pada 'new-tweet.rds'.

```
```{r}
api_key<- "8w1JX6MZjrI0OGKLl0ljvfwXu"
api_secret<- "yT2NkanLqicGid6P8oEHzMvwFC014FSzAilyGdb2BM4DOhf56E"
access_token<- "397904324-1j3PA9PrdrSUV7B4g8iK5J1gMdBlcdrGZpWi8f38"
access_token_secret<-
"sZL2yUSgFtTzWWmXBMzSsNLpew3QgV1FMDvqmUAQYGfT3"
setup_twitter_oauth(api_key,api_secret,access_token,access_token_se
cret)
```

```{r}
tw = searchTwitter('cryptocurrency',
                    n = 1000,
                    retryOnRateLimit = 10e5, lang = "id")
#retryOnRateLimit untuk looping
saveRDS(tw,file = 'new-tweet.rds')
```
```

**Listing Program 3.2** Menginputkan Api key Twitter dan Melakukan Search dengan caption Cryptocurrency

Pada listing program 3.3, kami melakukan read data yang tadi telah disimpan lalu dilanjutkan dengan proses Data Cleansing.

```
tw <- readRDS('new-tweet.rds')
```

```

d = twListToDF(tw) #convert twitter list to data
#menampilkan semua tweet yang kita mining
komen <- d$text
komenc <- Corpus(VectorSource(komen))

##hapus URL
removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
twitclean <- tm_map(komenc, removeURL)
##hapus New Line
removeNL <- function(y) gsub("\n", "", y)
twitclean <- tm_map(twitclean, removeNL)
##hapus koma
replacecomma <- function(y) gsub(",", "", y)
twitclean <- tm_map(twitclean, replacecomma)
##hapus retweet
removeRT <- function(y) gsub("RT ", "", y)
twitclean <- tm_map(twitclean, removeRT)
##hapus titik
removetitik2 <- function(y) gsub(":", "", y)
twitclean <- tm_map(twitclean, removetitik2)
##hapus titik koma
removetitikkoma <- function(y) gsub(";", " ", y)
twitclean <- tm_map(twitclean, removetitikkoma)
#hapus titik3
removetitik3 <- function(y) gsub("p.", "", y)
twitclean <- tm_map(twitclean, removetitik3)
#hapus &
removeamp <- function(y) gsub("&", "", y)
twitclean <- tm_map(twitclean, removeamp)
#hapus Mention
removeUN <- function(z) gsub("@\\w+", "", z)
twitclean <- tm_map(twitclean, removeUN)
#hapus space dll
remove.all <- function(xy) gsub("[^[:alpha:][:space:]]*", "", xy)
twitclean <-tm_map(twitclean,stripWhitespace)
inspect(twitclean[1:10])
twitclean <- tm_map(twitclean,remove.all)
twitclean <- tm_map(twitclean, removePunctuation) #tanda baca
twitclean <- tm_map(twitclean, tolower) #mengubah huruf kecil
myStopwords <- readLines("stop.txt", warn = FALSE)
twitclean <- tm_map(twitclean,removeWords,myStopwords)
twitclean <- tm_map(twitclean , removeWords,
 c('kalo','gak','org',''))

```

***Listing Program 3.3*** Proses Read Data dan Data Cleansing

Kemudian pada listing program 3.4 berfungsi untuk menghapus data yang hilang atau kosong. Sehingga data yang sudah benar-benar bersih akan disimpan pada file ‘tweet\_cryptocurrency.csv’.

```

#HAPUS DATA KOSONG
try.error = function(x)
{
 # create missing value
 y = NA
 # tryCatch error
 try_error = tryCatch(tolower(x), error=function(e) e)
 # if not an error

```

```

 if (!inherits(try_error, "error"))
 y = tolower(x)
 # result
 return(y)
 }
 # lower case using try.error with sapply
 twitclean = sapply(twitclean, try.error)
 # remove NAs in some_txt
 twitclean = twitclean[!is.na(twitclean)]
 names(twitclean) = NULL
  ```

  ```{r}
 # dataframe data yg sudah bersih
 dataframe<-data.frame(text=unlist(sapply(twitclean, `[`)),
 stringsAsFactors=F)
 View(dataframe)
 write.csv(dataframe, 'tweet_cryptocurrency.csv')

```

**Listing Program 3.4** Proses Menghapus Data Kosong dan Menyimpan Data yang telah Dibersihkan

Setelah itu kami melakukan sentimen analisis menggunakan Naïve Bayes untuk menghitung presentasi emosi dan menampilkan dalam bentuk Scatterplot.

```

Naive Bayes
library(e1071)
library(caret)
library(syuzhet)

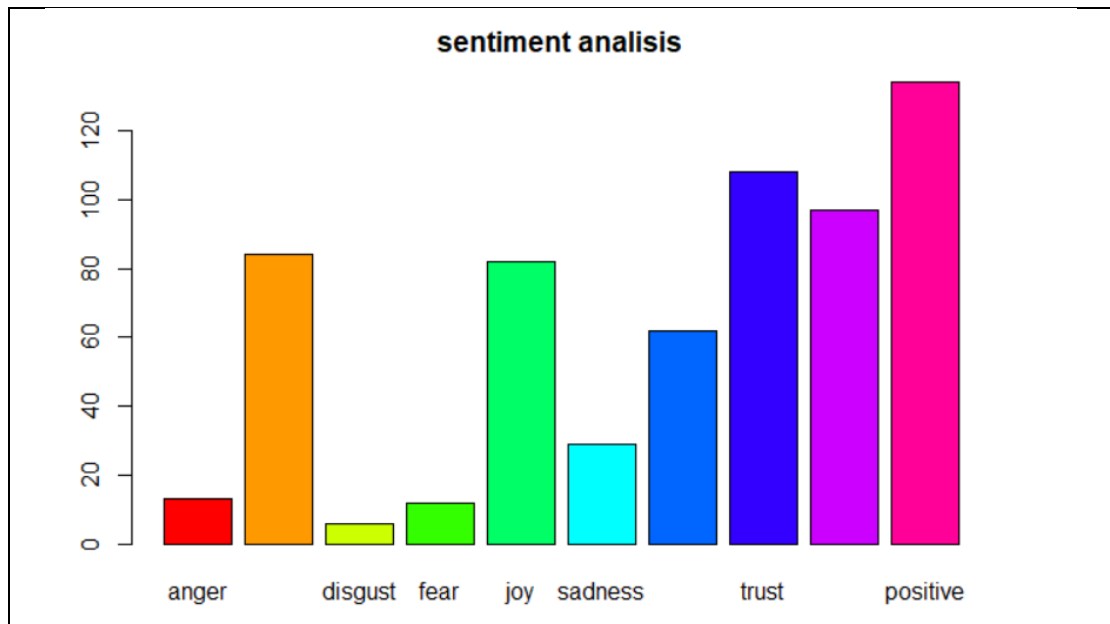
#digunakan untuk membaca file csv yang sudah di cleaning data
cryptocurrency_dataset <-
read.csv("tweet_cryptocurrency.csv", stringsAsFactors = FALSE)

#digunakan untuk mengeset variabel cloumn text menjadi char
review <- as.character(cryptocurrency_dataset$text)

#memanggil sentimen dictionary untuk menghitung presentasi dari
beberapa emotion dan mengubahnya ke dalam text file
get_nrc_sentiment('happy')
get_nrc_sentiment('excitement')
s<-get_nrc_sentiment(review)
review_combine<-cbind(cryptocurrency_dataset$text,s)
par(mar=rep(3,4))
barplot(colSums(s), col=rainbow(10), ylab='count', main='sentiment
analisis')

```

**Listing Program 3.5** Proses Analisis Menggunakan Naïve Bayes



**Gambar 3.1** Output Scatterplot

Selanjutnya pada listing program 3.6 yaitu menampilkan wordcloud dari data yang telah melewati Data Cleansing.

```
```{r}
## Wordcloud

library(tm) #library untuk penggunaan corpus dalam cleaning data
library(RTextTools) #library untuk penggunaan corpus dalam cleaning data
library(e1071) #library yang terdapat sebuah algoritma naivebayes
library(dplyr) #library yang terdapat sebuah algoritma naivebayes
library(caret) #library yang terdapat sebuah algoritma naivebayes
library(shiny) #library untuk memanggil shinny
library(vroom) #membersihkan data
library(here) #membersihkan data
library(ggplot2) #membuat barplot, grafik, dll
library(plotly)
library(syuzhet)

df<-read.csv("tweet_cryptocurrency.csv",stringsAsFactors = FALSE)
#membaca file CSV
glimpse(df) #melihat tipe dan struktur objek. Jika objek tersebut
adalah data frame

#Atur seed generator bilangan acak R, yang berguna untuk membuat
simulasi atau objek acak yang dapat direproduksi.
set.seed(20) #seed sebesar 20
df<-df[sample(nrow(df)),]
df<-df[sample(nrow(df)),]
glimpse(df) #melihat tipe data
df$X=as.factor(df$X) #mengubah menjadi faktor
#menampilkan semua tweet yang kita mining
corpus<-Corpus(VectorSource(df$text))
corpus
#melihat data yang telah i corpus
```



```

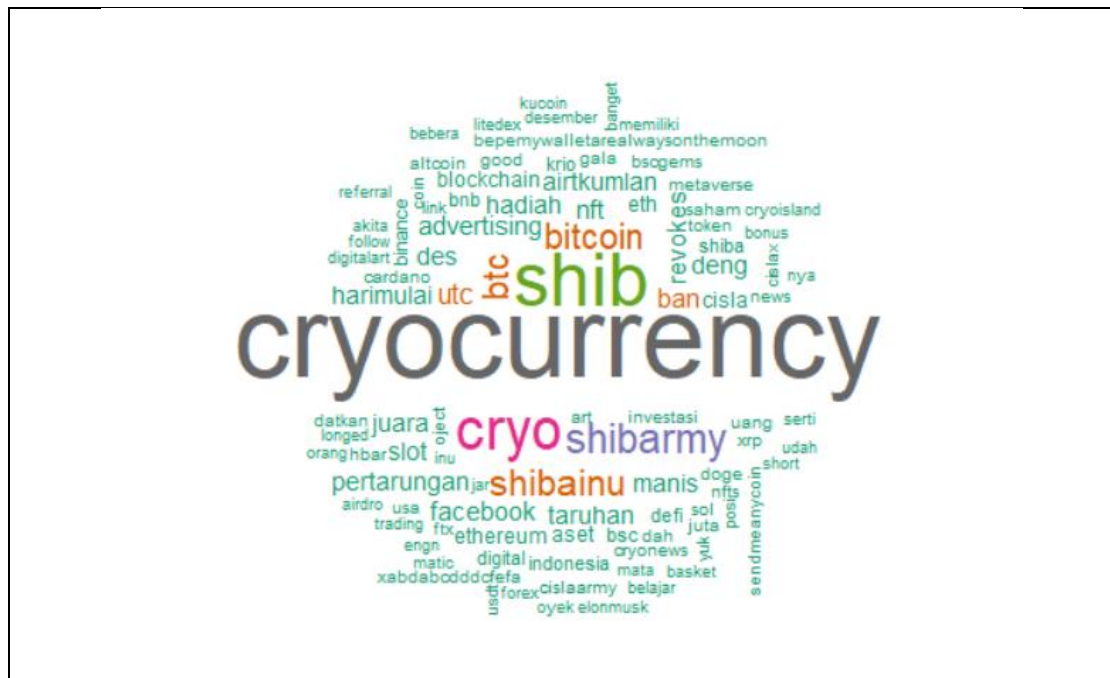
inspect(corpus[1:10])

#fungsinya untuk membersihkan data data yang tidak dibutuhkan,
menyaring data yang sbelumnya belum tersaring
corpus.clean<-corpus%>%
  tm_map(content_transformer(tolower))%>% #digunakan untuk mengubah
huruf besar dari string menjadi string huruf kecil
  tm_map(removePunctuation)%>% #menghapus anda baca
  tm_map(removeNumbers)%>% #menghapus nomor
  tm_map(removeWords, stopwords(kind="en"))%>% #menghapus stopwords
  tm_map(stripWhitespace)
dtm<-DocumentTermMatrix(corpus.clean)
inspect(dtm[1:10,1:20])
#klasifikasi dari data cryptocurrency dengan melakukan data
training dan data test
df.train<-df[1:589,]
df.test<-df[590:1177,]
dtm.train<-dtm[1:589,]
dtm.test<-dtm[590:1000,]
corpus.clean.train<-corpus.clean[1:589]
corpus.clean.test<-corpus.clean[590:1000]
#biasanya teks diubah menjadi Document-Term Matrix (DTM) melalui
proses yang bernama tokenization. Tokenization berfungsi memecah 1
kalimat menjadi beberapa term. term dapat berupa 1 kata, pasangan 2
kata, dan seterusnya.
dim(dtm.train)
fivefreq<-findFreqTerms(dtm.train,5) #frekuensi kemunculan kata
tersebut pada dokumen
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
}
#data train dan test naive bayes
trainNB<-apply(dtm.train.nb,2,convert_count)
testNB<-apply(dtm.test.nb,1,convert_count)
classifier<-naiveBayes(trainNB,df.train$X,laplace = 1)
library(wordcloud)
#menampilkan tampilan dari wordcloud
wordcloud(corpus.clean,min.freq =
4,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))
```

```

***Listing Program 3.6*** Proses Menampilkan Wordcloud



**Gambar 3.2** Output Wordcloud

Dan yang terakhir adalah menampilkan hasil data yang telah diproses tadi ke dalam aplikasi shiny.

```
Shiny
#membuka file csv
twitter <- read.csv(file="tweet_cryptocurrency.csv",header=TRUE)
#membuka text file pada data frame twitter
tweet <- twitter$text

bagian yang mengatur tampilan web, baik input maupun outpun yang
akan ditampilkan dalam web app.
ui <- fluidPage(
 titlePanel("Penggunaan Kata Cryptocurrency Pada Twitter"),
 #halaman judul dr fluid page
 mainPanel(#tab pada fluidpage
 #plot output : untuk scatterplot
 tabsetPanel(type = "tabs",
 tabPanel("Scatterplot", plotOutput("scatterplot")),
 #tab berupa scatterplot/grafik
 tabPanel("Data Twitter",
 DT::dataTableOutput('tbl')), #tab berupa data clening twitter
 tabPanel("Wordcloud", plotOutput("Wordcloud")) #tab
 berupa worldcloud
)
)
)

SERVER
Disinialah tempat dimana data akan dianalisis dan diproses lalu
hasilnya akan ditampilkan atau diplotkan pada bagian mainpanel() ui
yang telah dijelaskan sebelumnya.
server <- function(input, output) {
```

```

Output Data
output$tbl = DT::renderDataTable({
 DT::datatable(twitter, options = list(lengthChange = FALSE)) #
data akan ditampilkan dalam beberapa halaman.
})

#Barplot
output$scatterplot <- renderPlot({vaksin_dataset<-
read.csv("tweet_cryptocurrency.csv",stringsAsFactors = FALSE)
review <-as.character(cryptocurrency_dataset$text)
get_nrc_sentiment('happy')
get_nrc_sentiment('excitement')
s<-get_nrc_sentiment(review)
review_combine<-cbind(cryptocurrency_dataset$text,s)
par(mar=rep(3,4))
barplot(colSums(s),col=rainbow(10),ylab='count',main='sentiment
analisis')
}, height=400)

#WordCloud
output$Wordcloud <- renderPlot({
 set.seed(20)
df<-df[sample(nrow(df)),]
df<-df[sample(nrow(df)),]
glimpse(df)
df$X=as.factor(df$X)
corpus<-Corpus(VectorSource(df$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[1:589,]
df.test<-df[590:1177,]
dtm.train<-dtm[1:589,]
dtm.test<-dtm[590:1000,]
corpus.clean.train<-corpus.clean[1:589]
corpus.clean.test<-corpus.clean[590:1000]
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)

```

```

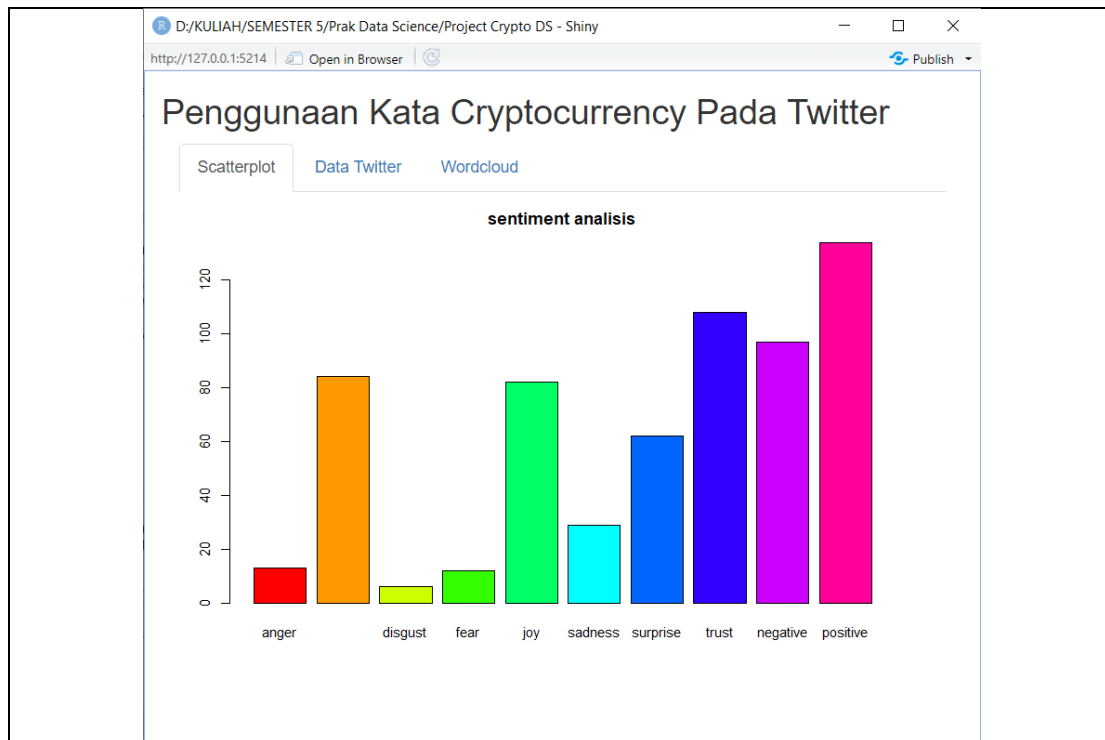
testNB<-apply(dtm.test.nb,1,convert_count)
classifier<-naiveBayes(trainNB,df.train$X,laplace = 1)
library(wordcloud)
wordcloud(corpus.clean,min.freq =
4,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))
})

}

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

```

Listing Program 3.7 Program Shiny



Gambar 3.3 Output Scatterplot pada Shiny



4. KESIMPULAN

Berdasarkan dari hasil analisis, perancangan dan pembahasan yang telah dilakukan, dapat diambil kesimpulan antara lain :

- 1) Progam Analisis Sentimen Kepopuleran Cryptocurrency dapat berjalan dengan baik, dapat menyeleksi data dari yang diambil langsung dari twitter, menghilangkan maupun mengganti karakter-karakter yang tidak dibutuhkan, dan menghasilkan analisis sentimen yang bernilai positif maupun negatif sesuai yang diharapkan.
- 2) Karena data diambil langsung dari api twitter sehingga akan terdapat perbedaan disetiap melakukan proses analisis sentimen karena data akan selalu *up to date*.