

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN TWITTER TENTANG
COVID-19 DI DUNIA MENGGUNAKAN *VADER*
*SENTIMENT ANALYSIS***



Muhammad Agil Suyuti 123200109

Rangga Restu Ramadhani 123200124

**PROGRAM STUDI INFORMATIKA
JURUSAN INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
YOGYAKARTA
2022**

HALAMAN PENGESAHAN

**ANALISIS SENTIMEN TWITTER TENTANG
COVID-19 DI DUNIA DENGAN METODE VADER**

Disusun Oleh:

Muhammad Agil Suyuti
Rangga Restu Ramadhani

123200109
123200124

Telah Diperiksa dan Disetujui oleh Asisten Praktikum pada
tanggal : 30 November 2022

Menyetujui
Asisten Praktikum

Asisten Praktikum



Dio Cahyo Putra S.Kom



Vicentius Willy Ardiyanto

1. PENDAHULUAN

Penyakit virus corona (COVID-19) adalah penyakit menular yang disebabkan oleh virus SARS-CoV-2. Sebagian besar orang yang tertular COVID-19 akan mengalami gejala ringan hingga sedang, dan akan pulih tanpa penanganan khusus. Namun, sebagian orang akan mengalami sakit parah dan memerlukan bantuan medis. Virus dapat menyebar dari mulut atau hidung orang yang terinfeksi melalui partikel cairan kecil ketika orang tersebut batuk, bersin, berbicara, bernyanyi, atau bernapas. Partikel ini dapat berupa droplet yang lebih besar dari saluran pernapasan hingga aerosol yang lebih kecil.

Covid-19 ini menyebabkan pandemi yang berawal pada tahun 2020, lebih tepatnya saat bulan Maret. Covid mencapai puncak penyebaran pada Juli 2021 dan Februari 2022, meskipun puncak penyebaran saat 2022 lebih tinggi sedikit, angka kematian nya sangat jauh berbeda dibandingkan saat 2021.

Dengan adanya pandemic Covid-19 ini banyak masyarakat yang melasa resah dan terkena dampak dari pandemi Covid-19, contohnya banyak tenaga kerja yang terkena PHK (pemutusan hubungan kerja), kemudian pada awal pandemic banyak kantor dan sekolah yang memerlukan adaptasi dari offline jadi online melalui zoom meeting. Karena tidak bisa kontak langsung antar manusia banyak usaha yang bangkrut dan kehilangan mata pencaharian, bisa dibilang efek dari Covid-19 merusak segala aspek seperti ekonomi, pendidikan, dll.

Oleh karena itu, kami ingin menganalisis tanggapan yang mereka eluhkan melalui media social twitter saat awal pandemi Covid-19 berlangsung. Kami memilih metode vader untuk menganalisis komen twitter karena Keuntungan menggunakan VADER polarity detection adalah tersedianya kamus yang berisi nilai dari setiap kata. Hasil Preprocess text akan di nilai berdasarkan lexicon apakah itu positif, negatif atau neutral dan menambahkan skor total (compound).

2. METODE

2.1 Pengambilan Data

Pada tahap pertama, yang dilakukan adalah proses pengambilan data dari *web* Kaggle. Sedangkan kata kunci pencarian yang digunakan adalah “COVID 19 Case Tweets”. Dataset yang didapat bernama “covid19_tweets.csv” yang di dalamnya berisi komentar warganet dalam bahasa Inggris. Dalam dataset tersebut, terdapat berbagai kolom yang kurang diperlukan dalam analisis sentimen, maka dari itu kami hanya mengambil kolom *text* dalam dataset tersebut. Total objek (baris) dalam dataset tersebut sebanyak 183930 baris yang nantinya akan kami saring dan tentukan berapa banyak baris yang akan kami gunakan.

'''{r} names(covid19_tweets)					
[1]	"user_name"	"user_location"	"user_description"	"user_created"	"user_followers"
[7]	"user_favourites"	"user_verified"	"date"	"text"	"user_friends"
[13]	"is_retweet"			"hashtags"	"source"

Gambar 2.1 Kolom yang dimiliki oleh dataset

Setelah data berhasil dikumpulkan, data tersebut akan disimpan ke dalam file dengan format csv yang nantinya akan dibaca dan dilakukan proses pada tahap selanjutnya.

2.2 Data Preprocessing

Data yang telah disimpan dalam sebuah *file* .csv dibaca dan disimpan dalam sebuah variabel untuk dapat dibersihkan hingga menjadi data matang yang dapat diolah dan dapat dilakukan analisis sentimen pada *refined data* tersebut.

Seperti *data cleaning* pada umumnya, kami menggunakan berbagai fungsi yang disediakan oleh *library(tm)* untuk melakukan *data cleaning*. Proses pembersihan data yang kami lakukan adalah dengan cara menghilangkan berbagai komponen yang tidak diperlukan dalam analisis senimen seperti menghapus: URL, *newline*, tanda baca (koma, titik dua, titik koma, dll.), menghapus username (@...), menghapus angka, mengubah huruf kapital menjadi *lower case*, dan menghilangkan *stopword* untuk menghapus kata-kata tidak berarti yang tidak diperlukan dalam melakukan analisis sentimen.

Setelah data sudah bersih, kami tentukan jumlah data yang akan kami analisis. Kami mengambil 1.000 baris dari 183.930 baris data. Dalam 1.000 data tersebut, kami saring lagi hingga menjadi total 707 baris data yang sudah benar-benar bersih dan siap untuk dianalisis sentimennya.

2.3 Vader Sentiment Analysis

Valence Aware Dictionary for sEntiment Reasoning, atau Vader merupakan algoritma NLP (*Natural Language Processing*) yang memadukan pendekatan sentimen *Lexicon* serta aturan tata bahasa dan konversi sintaksis untuk mengekspresikan polaritas dan intensitas sentimen. Vader adalah *open-sourced package* dalam NLTK (*Natural Language Toolkit*).

Tabel 2.1 Contoh pengaruh perbedaan karakter pada kamus Vader

Input	neg	neu	pos	compound
"This laptop is a good deal"	0	0.58	0.42	0.44
"This laptop is a very good deal"	0	0.61	0.39	0.49
"This laptop is a very good deal!!"	0	0.57	0.43	0.58
"This laptop is a very good deal!! :-)"	0	0.44	0.56	0.74
"This laptop is a VERY good deal!! :-)"	0	0.393	0.61	0.82

Untuk menghitung skor sentimental dari keseluruhan teks, Vader memindai seluruh teks berdasarkan fitur sentimen yang diketahui. Skor per teks diketahui berdasarkan intensitas dan polaritas sesuai aturan, lalu dilakukan normalisasi skor akhir menjadi (-1, 1) menggunakan fungsi:

$$\frac{x}{\sqrt[2]{x^2 + \alpha}}$$

Di mana x = jumlah skor valensi kata penyusun, dan simbol α didefinisikan sebagai konstanta normalisasi (*default value* = 15). Contoh pengimplementasian rumus :

weve progressed joking {0, 0, 0.9}	$\frac{0.9}{\sqrt[2]{0.9^2 + 15}} = 0,226$	Positive
students stressed due pandemic mentally prepare exams moreover {0, -1.4, 0, 0, 0, 0, 0}	$\frac{1.4}{\sqrt[2]{1.4^2 + 15}} = -0,339$	Negative
vaccine coverings stay feels massive shift towards {0, 0, 0, 0, 0, 0, 0, 0}	$\frac{0}{\sqrt[2]{0 + 15}} = 0$	Neutral

2.4 Pembuatan Model Menggunakan *app Shiny*

R Shiny merupakan sebuah package dalam bahasa pemrograman R yang digunakan untuk membangun aplikasi interaktif berbasis website. *R Shiny* ini merupakan gabungan antara keluaran komputasi statistic R dengan proses interaksi website modern. Dalam penelitian ini, *R Shiny* digunakan untuk menampilkan visualisasi data yang akan ditampilkan ke dalam 2 bentuk, yaitu *barplot* dan *wordcloud*.

Barplot merupakan diagram batang yang akan menampilkan perbandingan pada satu atau lebih variabel data. Data yang disajikan berupa grafik batang dalam bentuk persegi panjang dimana panjangnya menyesuaikan berdasarkan nilai dari hasil klasifikasi masing-masing variabel. Dengan *barplot*

ini, kita dapat mengetahui dengan cepat perbandingan nilai yang dimiliki tiap variabel dan memudahkan kita dalam mengambil kesimpulan nilai mana yang paling tinggi hingga nilai yang paling rendah.

Sedangkan *wordcloud* atau biasa dikenal dengan text cloud merupakan sebuah metode untuk menampilkan data teks secara visual. Dengan menggunakan *wordcloud* ini, data yang ditampilkan terlihat lebih menarik namun tetap informatif dan mudah dipahami. *Wordcloud* akan menampilkan satu kata dengan ukuran huruf yang lebih besar dibanding dengan lainnya untuk satu kata yang paling sering muncul atau frekuensi penggunaan katanya paling besar. Sedangkan ukuran huruf akan semakin mengecil jika frekuensi penggunaan kata tersebut semakin sedikit.

Adapun fitur lainnya yang kami munculkan pada R Shiny, yaitu data tabel berisi sentimen dan Frequency of Words yang juga ditampilkan dalam bentuk *barplot*.

3. HASIL DAN PEMBAHASAN

3.1 *Import Library yang Dibutuhkan*

```
library(syuzhet)      # package analisis sentimen
library(RTextTools)   # melakukan klasifikasi text
library(e1071)        # melakukan analisa statistik
library(tidyverse)    # merapikan data
library(tm)           # membersihkan data
library(textdata)     # download dan load berbagai dataset
library(ggplot2)      # membuat plot
library(dplyr)        # manipulasi data
library(plyr)         # membagi dan menggabungkan data
library(shiny)        # membuat UI
library(wordcloud)    # membuat world cloud
library(RColorBrewer) # menggunakan pallet warna
library(corpus)       # training klasifikasi dan regresi
library(vader)        # melakukan analisis sentimen
library(markdown)     # melakukan format data
library(DT)           # memberikan output data pada tabel
library(caret)        # melakukan klasifikasi data
library(SnowballC)    # mengubah kata menjadi kata dasar
```

Listing 3.1 *Import library*

3.2 *Pengambilan Data*

```
tweetCovid <- covid19_tweets$text
covid_corpus <- Corpus(VectorSource(tweetCovid))
```

Listing 3.2 *Mengambil data berupa teks dari dataset*

3.3 *Data Preprocessing*

```
#menghapus URL
removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
tweetclean <- tm_map(covid_corpus, removeURL)

#menghapus NewLine
removeNL <- function(y) gsub("\n", " ", y)
tweetclean <- tm_map(tweetclean, removeNL)

#menghapus koma
replacecomma <- function(y) gsub(",", "", y)
tweetclean <- tm_map(tweetclean, replacecomma)

#menghapus titik2
removetitik2 <- function(y) gsub(":", "", y)
tweetclean <- tm_map(tweetclean, removetitik2)

#menghapus titik koma
removetitikkoma <- function(y) gsub(";", " ", y)
tweetclean <- tm_map(tweetclean, removetitikkoma)

#menghapus titik3
removetitik3 <- function(y) gsub("p...", "", y)
tweetclean <- tm_map(tweetclean, removetitik3)

#menghapus RT(retweet)
removeRT <- function(y) gsub("RT", "", y)
tweetclean <- tm_map(tweetclean, removeRT)

#menghapus &
```

```

removeamp <- function(y) gsub("&", "", y)
tweetclean <- tm_map(tweetclean, removeamp)

#menghapus username
removeUN <- function(z) gsub("@\\w+", "", z)
tweetclean <- tm_map(tweetclean, removeUN)

#menghapus space dan lainnya
remove.all <- function(xy) gsub("[^:alpha:][:space:]]*", "", xy)
tweetclean <- tm_map(tweetclean, remove.all)

#menghapus tanda baca
tweetclean <- tm_map(tweetclean, removePunctuation)

#menghilangkan angka
tweetclean <- tm_map(tweetclean, removeNumbers)

#menghilangkan white space
tweetclean <- tm_map(tweetclean, stripWhitespace)

#mengubah ke huruf kecil
tweetclean <- tm_map(tweetclean, tolower)

#menghilangkan stopwords
stopwords = readLines("stopwords.txt")
tweetclean <- tm_map(tweetclean, removeWords, stopwords)

```

Listing 3.3 *data cleaning*

3.4 Seleksi Data

```

dataframe<-data.frame(text=unlist(sapply(tweetclean, `[`)),
stringsAsFactors=F)
dataframe_decreased<-head(dataframe, n=1000)

write.csv(dataframe_decreased, file = 'tweetclean_covid.csv')

```

Listing 3.4 *data filtering*

```

#Membaca file csv yang sudah menjalani proses cleaning data
covid_fulldata <-
read.csv("tweetclean_covid.csv",stringsAsFactors = FALSE)
covid_data <- covid_fulldata$text

#Menghapus data yang hilang
sorted_df <- sort(covid_data, decreasing = TRUE)
covid_data <- sorted_df[-(748:1000)]

#Menghapus data duplikasi
covid_data <- covid_data[!duplicated(covid_data)]

```

Listing 3.5 *Pembersihan data dari data duplikasi dan data kosong*

3.5 Implementasi *Vader Analysis Sentiment*

```

#Vader Sentiment Analysis
library(tidyr)

df = data.frame(covid_data)
colnames(df) <- c('Tweet')

```



```

data_tweet = df

df_temp = vader_df(df['Tweet'])
df["positive"]<- df_temp['pos']
df["negative"] <- df_temp['neg']
df["neutral"] <- df_temp['neu']
df["compound"] = df_temp['compound']

df['label'] <- NA

for(i in 1:nrow(df)){
  if (df[i,"compound"] >= 0.05){
    df[i,"label"] = "Positive"
  } else if (df[i,"compound"] < 0 ){
    df[i,"label"] = "Negative"
  } else {
    df[i,"label"] = "Netral"
  }
}

df_classified = data.frame(df["Tweet"], df["label"])
view(df_classified)

sentiment count = count(df_classified$label)

```

Listing 3.6 Analisis sentimen beserta pemberian label

3.6 Corpus Peparation

```

#Corpus Cleanse
corpus <- Corpus(VectorSource(data_tweet))
corpus <- tm_map(corpus, removeWords, c(stopwords(),
"covid"))

removw_url <- function(x) gsub("http[^[:space:]]*", "", x)
corpus <- tm_map(corpus, content_transformer(removw_url))

removeNumPunct <- function(x)
gsub("[^[:alpha:][:space:]]*", "", x)
corpus <- tm_map(corpus, content_transformer(removeNumPunct))
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, content_transformer(tolower))
corpus <- tm_map(corpus, stripWhitespace)
corpus <- tm_map(corpus, stemDocument)

tdm <- TermDocumentMatrix(corpus)
tdm <- removeSparseTerms(tdm, sparse = 0.98)
tdm <- as.matrix(tdm)
w = sort(rowSums(tdm), decreasing = T)

```

Listing 3.7 Pembersihan data untuk diambil data per kata

3.7 Shiny App

```

library(markdown)
library(DT)

ui <- fluidPage(
  titlePanel("Tweets Kasus Covid-19 di Dunia"), #halaman
  judul
  mainPanel(

```

```

        tabsetPanel(type = "tabs",
                    tabPanel("Data Twitter",
DT::dataTableOutput('data')),
                    tabPanel("Sentiment Analysis",
DT::dataTableOutput('sentiment')),
                    tabPanel("Scatterplot",
plotOutput('scatterplot')),
                    tabPanel("Frequency Words",
plotOutput('freqword')),
                    tabPanel("Wordcloud",
plotOutput('wordcloud'))
        )
    )
)

```

Listing 3.8 *Set up Shiny (server)*

```

server <- function(input, output) {
  output$data <- DT::renderDataTable({
    DT::datatable(df["Tweet"], options = list(lengthChange =
FALSE))
  })

  output$sentiment <- DT::renderDataTable({
    DT::datatable(df_classified, options = list(lengthChange
= FALSE))
  })

  output$scatterplot <- renderPlot({
    barplot(sentiment_count$freq,
            names.arg = c("Negative", "Neutral", "Positive"),
            main='Sentiment Analysis',
            col = rainbow(3)
    )
  })

  output$freqword<- renderPlot({
    barplot(w[c(3,4,5,7,8)],
            las=2,
            main = "Frequency of Words",
            col= rainbow(5))
  })

  output$wordcloud<- renderPlot({
    wordcloud(corpus, min.freq = 3,
              max.words=100, random.order=FALSE, rot.per=0.40,
              colors=brewer.pal(8, "Dark2"))
  })
}

```

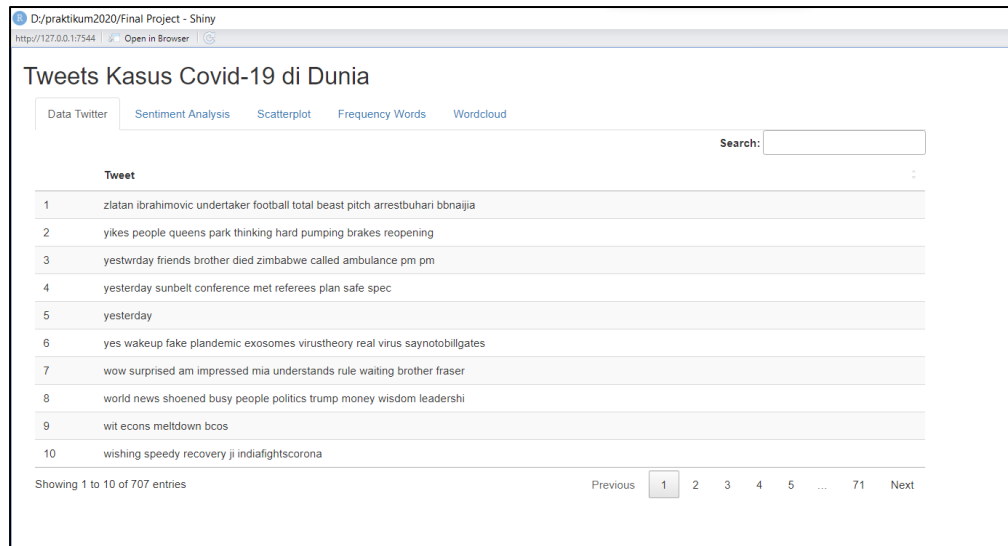
Listing 3.9 *Set up Shiny (UI)*

```

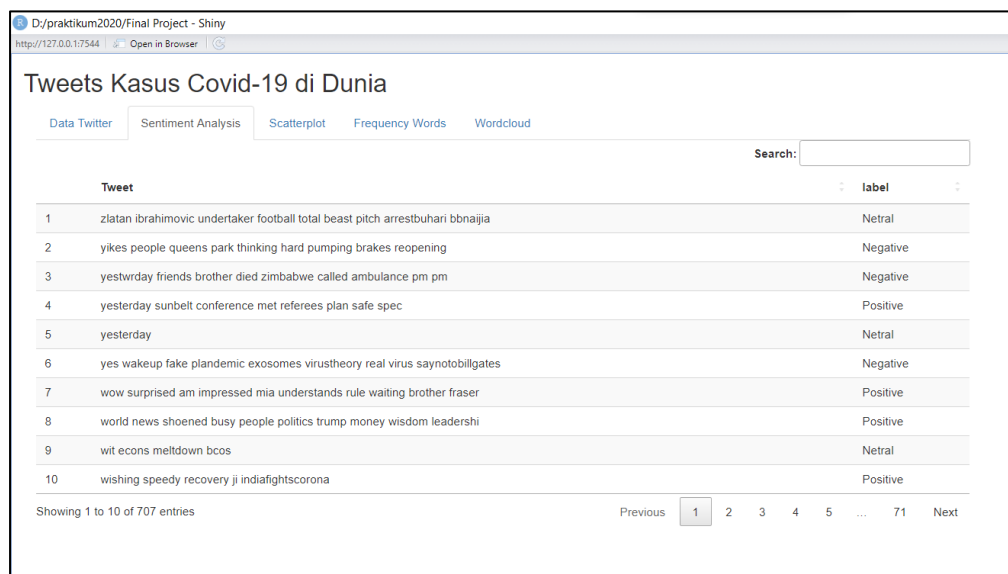
shinyApp(ui = ui, server = server)

```

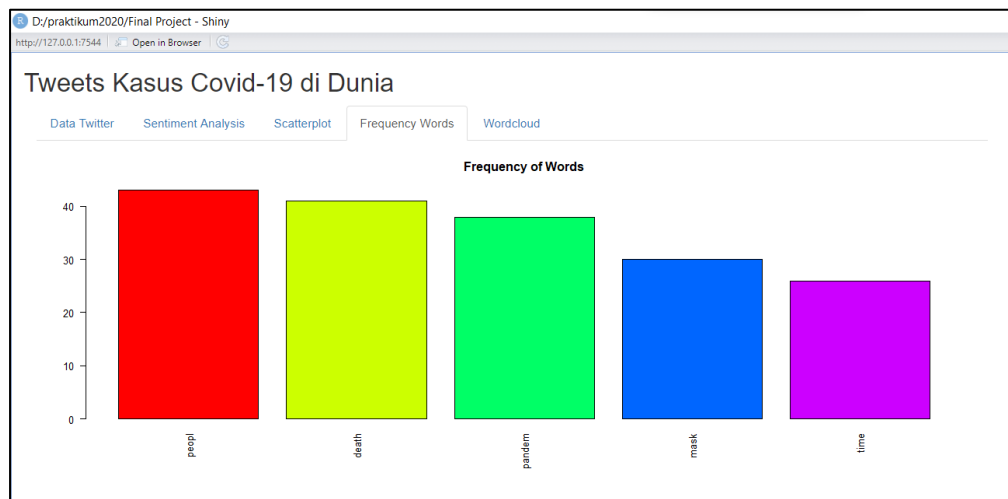
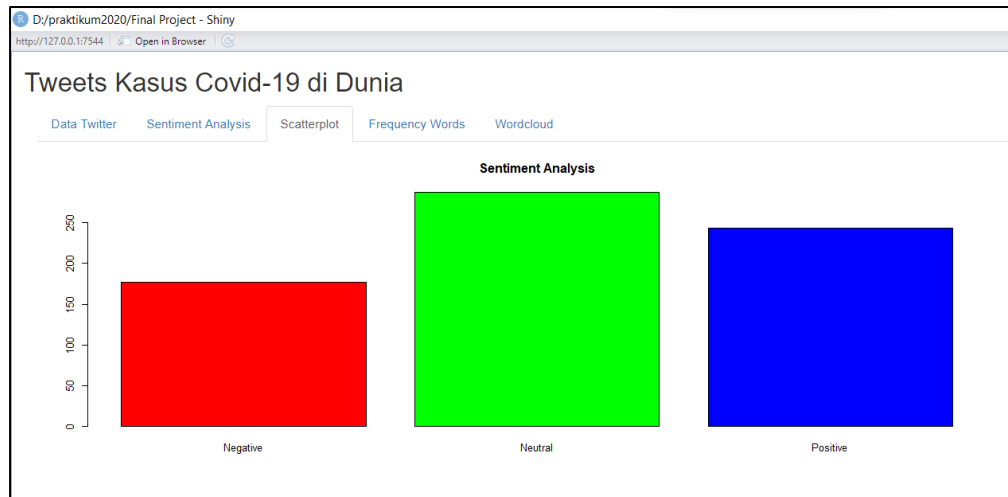
Listing 3.10 *Run shiny app*



Gambar 3.1 Tampilan utama berisi data *Twitter*



Gambar 3.2 Tampilan analisis sentimen



4. KESIMPULAN

Kesimpulan yang kami adalah pandemi Covid-19 ini menimbulkan pro dan kontra , ada yang menyalahkan pemerintah dan orang lain, ada juga yang mengikuti aturan pemerintah tanpa bertanya karena tau yang dilakukan pemerintah yang terbaik untuk masyarakat.

Dari analisis kami tanggapan masyarakat di twitter lebih ke arah positif. Orang-orang di twitter lebih ke mengajak orang-orang untuk memakai masker, menjaga jarak, tes Covid-19 di fasilitas kesehatan terdekat, tetap dirumah, khawatirkan kesehatan orang yang sudah tua disekitar anda karena lebih rawan terjangkit, dan segera melapor jika anda melihat atau merasakan gejala Covid-19

Hanya sedikit tanggapan masyarakat yang bersifat negatif saat awal pandemi ini.