

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE**

**Analisis Sentimen Terkait Topik Bepergian di Masa
Pandemi pada Sosial Media Twitter**



Vincentius Willy Ardiyanto 123190100
Putri Fajriana 123190110

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

1. PENDAHULUAN

Pada awal tahun 2020, masyarakat Indonesia dihebohkan dengan masuknya virus corona atau biasa disebut dengan Covid-19. Masuknya virus ini mengakibatkan adanya pembatasan ruang gerak dan aktivitas masyarakat di luar rumah. Dengan kata lain, untuk mencegah penyebaran, saat ini pemerintah menetapkan adanya PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat). PPKM sendiri terdiri dari beberapa level dimana semakin kecil levelnya, maka semakin longgar pula ruang gerak masyarakat di luar rumah. Berdasarkan jumlah kasus Covid-19 dan status vaksinasi, kini pemerintah menetapkan status PPKM berada di level rendah.

Banyak tempat wisata, hotel, dan tempat lainnya yang sudah mulai beroperasi diiringi dengan banyaknya masyarakat yang bepergian untuk melepas rasa jenuhnya. Namun, pendapat masyarakat mengenai bepergian di masa pandemi saat ini cukup beragam. Twitter menjadi salah satu situs jejaring sosial yang banyak digunakan untuk membagikan pendapat maupun informasi salah satunya yaitu terkait bepergian di masa pandemi sekarang.

Setiap pendapat yang ditulis melalui twitter dapat memiliki nilai positif, negatif, atau tidak keduanya (netral). Sehingga diperlukan sebuah proses analisis sentiment untuk mengklasifikasikan suatu pendapat yang ditulis. Menurut Artanti et al. (2018), analisis sentiment sendiri merupakan sebuah proses menemukan pendapat pengguna terhadap beberapa topik yang disampaikan oleh pengguna dan menentukan apakah pendapat itu bernilai positif, negatif, atau netral.

Tujuan dari penelitian ini adalah melakukan pengumpulan data twitter secara real time yang diperlukan untuk analisis sentimen berkaitan dengan topik bepergian di masa pandemi. Data yang didapat dari TwitterAPI akan diproses dan dilakukan pembersihan data untuk memilah data sesuai dengan yang diinginkan menggunakan Bahasa pemrograman R hingga dapat dilakukan analisis dan visualisasinya.

2. METODE

2.1 Pengambilan Data

Pada tahap pertama, yang dilakukan adalah proses pengambilan data dari TwitterAPI dimana diperlukan 4 key agar bisa mengaksesnya yaitu *consumer key*, *consumer secret key*, *access token key*, dan *access token secret key*. Sedangkan kata kunci pencarian yang digunakan adalah *pandemic* dan *travelling* dengan pengaturan *tweet* menggunakan Bahasa Inggris. 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 Preprocessing

Setelah mendapatkan data kemudian dilakukan pembersihan data dengan berbagai indikator. Hal ini dilakukan untuk mengekstraksi data set yang tidak terstruktur agar tidak menyulitkan dalam proses analisis sentimen nantinya. Beberapa hal yang dilakukan dalam proses pembersihan data adalah:

- a. *Remove Punctuation*, yaitu *cleaning* data dengan menghilangkan tanda baca atau karakter *non alphabet* seperti simbol, spasi, dan lain-lain.
- b. *Case Folding*, yaitu proses mengubah setiap kata menjadi sama. Dalam penelitian ini digunakan fungsi *lowercase* untuk menghilangkan huruf kapital atau mengubah huruf menjadi kecil semua
- c. *Remove Number*, yaitu proses menghapus angka. Karena pada umumnya penulisan angka secara berulang merupakan hal yang salah sehingga apabila menemukan kata yang menggunakan tambahan angka maka perlu dihapus.
- d. *Remove URL*, yaitu proses menghapus URL karena kemunculan sebuah url dari data twitter membuat data tidak efektif dan tidak memiliki arti.
- e. *Remove Emoticon*, yaitu proses menghilangkan emoticon karena data yang diperlukan hanya berupa string.
- f. *Remove Stopword*, untuk menghapus kata yang kurang penting dan bersifat terlalu umum seperti kata yang frekuensi kemunculannya cukup banyak dibandingkan dengan kata lainnya seperti kalian, aku, dan lain-lain.

2.3 Visualisasi Analisis Sentimen dengan R 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 statistik R dengan proses interaksi website modern. Dalam penelitian ini, *R Shiny* digunakan untuk menampilkan visualisasi data yang akan ditampilkan ke dalam dua bentuk, yaitu *barplot* dan *wordcloud*.

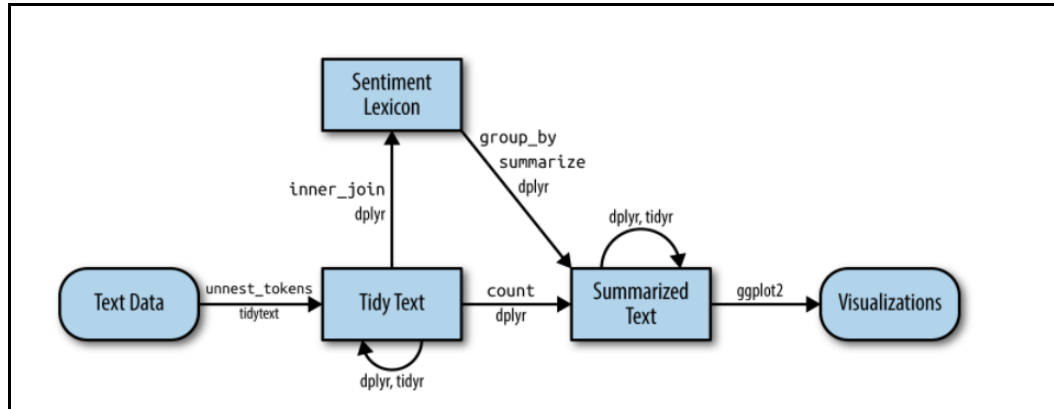
Barplot merupakan diagram batang yang akan menampilkan perbandingan pada satu atau lebih variabel data. Data yang disajikan berupa grafik batang dalam bentuk persegi panjang dimana panjangnya menyesuaikan berdasarkan nilai dari hasil klasifikasi masing-masing variabel. Dengan *barplot* ini, kita dapat

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.

2.4 Sentimen Analisis dengan Tidytext

Sentimen analisis dengan *tidytext* ini memungkinkan kita untuk menganalisis kata mana yang paling sering digunakan dalam dokumen (*frequent word*) dan untuk membandingkan dokumen, tetapi sekarang mari kita selidiki topik yang berbeda. Ketika pembaca manusia mendekati sebuah teks, kita menggunakan pemahaman kita tentang makna emosional kata-kata untuk menyimpulkan apakah suatu bagian teks itu positif, netral atau negatif, atau mungkin dicirikan oleh emosi lain yang lebih bernuansa seperti takut, terkejut atau jijik. Kita dapat menggunakan alat penambangan teks untuk mendekati konten emosional teks secara terprogram, seperti yang ditunjukkan pada figur di bawah.



Figur 2.1 Diagram alur tidytext untuk analisis sentimen

Ada berbagai metode dan kamus yang ada untuk mengevaluasi pendapat atau emosi dalam teks. *Package tidytext* menyediakan akses ke beberapa leksikon sentimen. Tiga leksikon tujuan umum adalah:

- AFINN dari Finn rup Nielsen,
- bing dari Bing Liu dan kolaborator, dan
- nrc dari Saif Mohammad dan Peter Turney.

Ketiga leksikon ini didasarkan pada unigram, yaitu, kata tunggal. Leksikon ini mengandung banyak kata bahasa Inggris dan kata-kata tersebut diberi skor untuk sentimen positif/negatif, dan juga mungkin emosi seperti kegembiraan, kemarahan, kesedihan, dan sebagainya. Leksikon nrc mengkategorikan kata-kata dalam mode

biner ("ya"/"tidak") ke dalam kategori positif, negatif, kemarahan, antisipasi, jijik, ketakutan, kegembiraan, kesedihan, kejutan, dan kepercayaan. Leksikon Bing mengkategorikan kata-kata dalam mode biner ke dalam kategori positif dan negatif. Leksikon AFINN memberikan kata-kata dengan skor yang berkisar antara -5 dan 5, dengan skor negatif menunjukkan sentimen negatif dan skor positif menunjukkan sentimen positif.

Tidak semua kata bahasa Inggris ada dalam leksikon karena banyak kata bahasa Inggris yang cukup netral. Penting untuk diingat bahwa metode ini tidak memperhitungkan kualifikasi sebelum kata, seperti dalam "tidak baik" atau "tidak benar"; metode berbasis leksikon seperti ini hanya didasarkan pada unigram. Untuk banyak jenis teks, tidak ada bagian sarkasme yang berkelanjutan atau teks yang dinegasikan, jadi ini bukan efek yang penting. Selain itu, kita dapat menggunakan pendekatan teks rapi untuk mulai memahami jenis kata negasi apa yang penting dalam teks tertentu.

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil dan pembahasan dari langkah analisis sentimen terkait “Topik Bepergian di Masa Pandemi pada Sosial Media Twitter” yang dilakukan.

3.1 Pengambilan Data

Data yang diambil menggunakan Twitter API disimpan ke dalam file dengan format csv dengan menggunakan teknik *scraping* data yang merupakan teknik mengambil atau mengekstrak data tertentu dari suatu sumber data. Sumber data yang dipakai di penelitian ini yaitu dari TwitterAPI dimana diperlukan *library* *rtweet* untuk mencari *tweet* dan twitter untuk autentikasi ke twitter. Kemudian untuk memperbaiki error pada library *rtweet* yaitu dengan menginstall *packages* *devtools*. Selanjutnya, ada empat *key* atau kunci yang dibutuhkan agar dapat mengakses data dari TwitterAPI dan keywords yang dicari dalam pencarian data yaitu *pandemic* dan *travelling* sebanyak 500 data dengan menggunakan pengaturan Bahasa Inggris.

```
library(rtweet)
library(twitter)
```

Listing Program 3.1 *Library* untuk pengambilan data

```
install.packages("devtools")
devtools::install_github("mkearney/rtweet")
```

Listing Program 3.2 *Install packages* “devtools”

```
setup_twitter_oauth(consumer_key = "IM0B6X2N1rYCOP7PmjahJXY5m",
                    consumer_secret =
                    "N9MpJxggVPfIi6CTh6lHwyAmh4aypkDESPetsAnld70WlfcUoo",
                    access_token = "1020303028769906688-
k8qcHgIfPgMPoPqwsRFWdNUoCqmPcp",
                    access_secret =
                    "znCQLuvNZ3JZQuzar86b98tAPSh8A2Fu5Qa0LjOI1SPmQ")
```

Listing Program 3.3 *Key / parameter* autentikasi

```
auth_setup_default()
travel_tweets <- search_tweets("pandemic+travelling -
filter:retweets", n=500, lang="en")
write.csv(travel_tweets, file = 'tweet_df.csv')
```

Listing Program 3.4 Autentikasi ke halaman twitter

3.2 Preprocessing

Cleaning atau pembersihan data *twitter* dilakukan dengan menghilangkan beberapa elemen yang tidak diperlukan seperti tanda baca, angka, *white space*, url, *emoticon*, dan lainnya menggunakan bahasa pemrograman R. Dalam proses pembersihan data, dibutuhkan beberapa *library* yang tersedia dalam R, yaitu *purrr* sebagai alternatif lain untuk melakukan *looping* (selain menggunakan *for loop*, *while loop*, atau *apply loop*) dengan memanfaatkan *map function*, *tidytext* untuk *text mining* maupun *preprocessing*, *dplyr* digunakan untuk memanipulasi data

dengan nama fungsi dan output yang konsisten, `tidyr` untuk merapikan suatu data atau membuat *pivot table* dari data, dan beberapa *library* lainnya seperti `textdata` dan `tm`. Sebelum data melalui proses pembersihan, bentuk data perlu diubah terlebih dahulu dari data frame menjadi data corpus. Setelah dilakukan proses pembersihan, data akan disimpan ke dalam data frame baru dan kemudian data frame tersebut diubah ke dalam bentuk file csv bernama `tweetclean_df`.

```
library(dplyr)
library(tidyr)
library(tidytext)
library(textdata)
library(purrr)
library(tm)
```

Listing Program 3.5 *Library* untuk pembersihan data

```
travellingTweets <- travel_tweets$text
travel_corpus <- Corpus(VectorSource(travellingTweets))
```

Listing Program 3.6 Mengubah data menjadi bentuk corpus

```
#menghilangkan tanda baca
tweet_clean <- tm_map(travel_corpus, removePunctuation)

#menghilangkan kapital
tweet_clean <- tm_map(tweet_clean, content_transformer(tolower))

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

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

#menghilangkan url
removeURL <- function(removeURL) gsub("http[^[:space:]]*", "",
removeURL)
tweet_clean <- tm_map(tweet_clean, removeURL)

#menghilangkan emoticon
removeEmoticon <- function(removeEmoticon) gsub("[^\x01-\x7F]",
"", removeEmoticon)
tweet_clean <- tm_map(tweet_clean, removeEmoticon)

#menghilangkan stopwords
setwd('D:\\Praktikum Data Science\\Final Project')
stopwords = readLines("stopwords.txt")
tweet_clean <- tm_map(tweet_clean, removeWords, stopwords)
```

Listing Program 3.7 Proses pembersihan data

```
#Menyimpan hasil pembersihan ke df baru
tweet_clean_df<-data.frame(text=unlist(sapply(tweet_clean,
`[`)), stringsAsFactors=F)

#mengubah df menjadi file csv
write.csv(tweet_clean_df,file = 'tweetclean_df.csv')
```

Listing Program 3.8 Menyimpan hasil pembersihan data

3.3 Visualisasi Analisis Sentimen dengan R Shiny

Untuk dapat memahami data yang sudah diolah dengan lebih mudah sebaiknya dilakukan visualisasi data. Pada bahasa R terdapat *library Shiny* yang bermanfaat untuk membuat aplikasi web sederhana dalam hal ini digunakan untuk platform visualisasi data/dashboard.

```
library(shiny)
library(shinydashboard)
library(shinythemes)
library(ggplot2)
library(plotly)
library(here)
library(vroom)
library(dplyr)
library(tidyverse)
library(tm)
library(wordcloud)
library(RTextTools)
library(syuzhet)
library(tidymodels)
library(textdata)
library(tidytext)
library(forcats)
```

Listing Program 3.9 Library yang digunakan untuk Shiny

```
ui <- fluidPage(
  theme = shinytheme("darkly"),
  titlePanel( h1("Tweet about Travelling During Pandemic
Situation")),
  mainPanel(
    tabsetPanel(type = "tabs",
      tabPanel("Tweet Data",
DT::dataTableOutput('data')), # Output Data Dalam Tabel
      tabPanel("NRC Sentiment Analysis",
plotOutput("nrc")), # Plot
      tabPanel("Bing Sentiment Analysis",
plotOutput("bing")), # Plot
      tabPanel("Frequent Word", plotOutput('freq')), #
Plot
```

Listing Program 3.10 Komponen UI


```

        tabPanel("Wordcloud", plotOutput("Wordcloud")) #
Plot Wordcloud
    ),
    checkboxInput(
        inputId = "themeToggle",
        label = icon("sun")
    )
)

```

Listing Program 3.11 Lanjutan komponen UI

```

# Output Data tabel
output$data = DT::renderDataTable({

    DT::datatable(data, options = list(lengthChange = FALSE))
})

```

Listing Program 3.12 Komponen Server – data

```

# Analisis sentimen NRC
output$nrc <- renderPlot({
    tweetsentiment<-cbind(tweetclean$text,sentiment)
    par(mar=rep(3,4))

    barplot(colSums(sentiment),col=rainbow(5),ylab='count',main='Ana
    lisis Sentimen NRC')
    }, height=400)

```

Listing Program 3.13 Komponen Server – NRC

```

#Analisis sentimen bing
output$bing <- renderPlot({

    separated <- data %>%
        filter(text != "nan") %>%
        unnest_tokens(word, text) %>%
        anti_join(stop_words)

    bing_word_counts <- separated %>%
        inner_join(get_sentiments("bing")) %>%
        count(word, sentiment, sort = TRUE) %>%
        ungroup()
        bing_word_counts %>%
        group_by(sentiment) %>%
        slice_max(n, n=10) %>%
        ungroup() %>%
        mutate(word = reorder(word,n)) %>%
        ggplot(aes(n,word,fill=sentiment))+
        geom_col(show.legend = FALSE) +
        facet_wrap(~sentiment, scales = "free_y") +
        labs(x = "Contribution to sentiment",
            y = NULL)
    }, height=400)

```

Listing Program 3.14 Komponen Server – bing

```
#Output kata yang sering muncul
output$freq <- renderPlot({
  corpus <- Corpus(VectorSource(tweetclean$text))
  corpus<-tm_map(corpus, removeWords,
c("travelling", "pandemic", "travel"))

  tdm<-TermDocumentMatrix(corpus)
  m<-as.matrix(tdm)
  v<-sort(rowSums(m),decreasing = TRUE)
  d<-data.frame(word=names(v),freq=v)
  barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
    main = "Frequent words",
    ylab = "Frequency")
})
```

Listing Program 3.15 Komponen Server – *Frequent Word*

```
#wordcloud
output$Wordcloud <- renderPlot({

  corpus <- Corpus(VectorSource(tweetclean$text))
  corpus<-tm_map(corpus, removeWords,
c("travelling", "pandemic", "travel"))

  wordcloud(corpus, random.order = F, max.words=100 ,
col=rainbow(100))
})
}
shinyApp(ui = ui, server = server, options = list(height =
"1080px"))
```

Listing Program 3.16 Komponen Server – *Wordcloud*

Berikut ini adalah hasil visualisasi menggunakan aplikasi web sederhana Shiny.

	text	sentiment
1	learnt travelling south korea pandemic	negative
2	bighitter read idiot travelling pandemic shut fuck lol	negative
3	tfw rail train travelling birmingham holyhead carriages jammed people toilets global pandemic strain spreading poor service dangerously irresponsible	negative
4	mattgallowaycbc call thick travelling pandemic leaders exacting example taking stringent steps halt spread closing borders instead implanting useless amp chaotic rules guess leadership	negative
5	cellnegounder testing requirements travelling pandemic incubation period short people tested negative hours time boarded flight test positive celinegounder	negative
6	booshmeister oh dont mid size suvs set itpickup truckgt yeah im travelling cargo rain compact suv jetta reserved month advance pre pandemic	neutral
7	bd sabcnews enca hey cyril job vax salesman travelling tax payers money peddling pandemic vax phuktard diktator mates	neutral
8	byoung wholemarsblog omicron spread vaccinated travelling individuals countries don allow unvaccinated travel narrative unvaccinated causing pandemic stop society digests	neutral
9	travelling pandemic fcking simpathy	negative
10	telegraphtravel travelling pandemic fcking simpathy	negative

Gambar 3.1 Tweet Data - Darkly

https://127.0.0.1:7279 Open in Browser

Tweet about Travelling During Pandemic Situation

[Tweet Data](#)
[NRC Sentiment Analysis](#)
[Bing Sentiment Analysis](#)
[Frequent Word](#)
[Wordcloud](#)

Search:

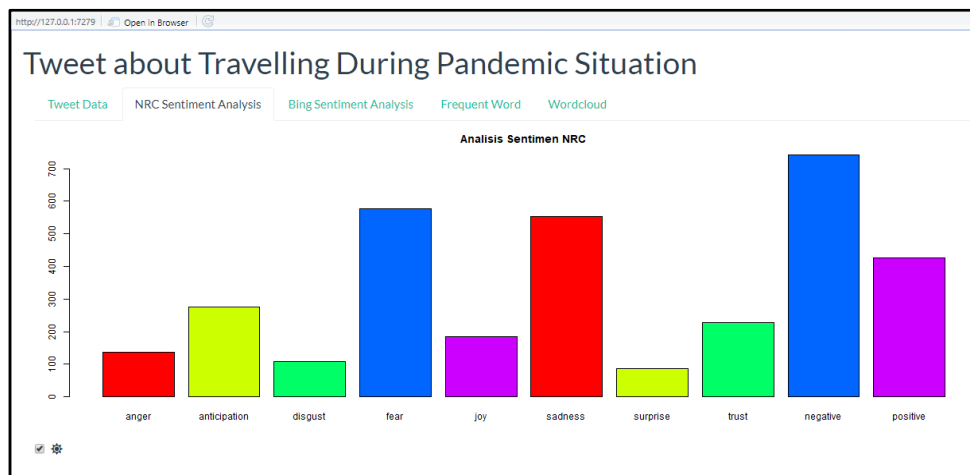
	text	sentiment
1	learnt travelling south korea pandemic	negative
2	bighitter read idiot travelling pandemic shut fuck lol	negative
3	tfw rail train travelling birmingham holyhead carriages jammed people toilets global pandemic strain spreading poor service dangerously irresponsible	negative
4	mattgallowaycbc call thick travelling pandemic leaders exacting example taking stringent steps halt spread closing borders instead implanting useless amp chaotic rules guess leadership	negative
5	celinegounder testing requirements travelling pandemic incubation period short people tested negative hours time boarded flight test positive celinegounder	negative
6	boosmeister oh dont mid size suvs set ltpickup truckgt yeah im travelling cargo rain compact suv jetta reserved month advance pre pandemic	neutral
7	bd sabcnws enca hey cyril job vax salesman travelling tax payers money peddling pandemic vax phuktard dictator mates	neutral
8	byoung wholemarsblog omicron spread vaccinated travelling individuals countries don allow unvaccinated travel narrative unvaccinated causing pandemic stop society digests	neutral
9	travelling pandemic fcking simpathy	negative
10	telegraphtravel travelling pandemic fcking simpathy	negative

Showing 1 to 10 of 401 entries

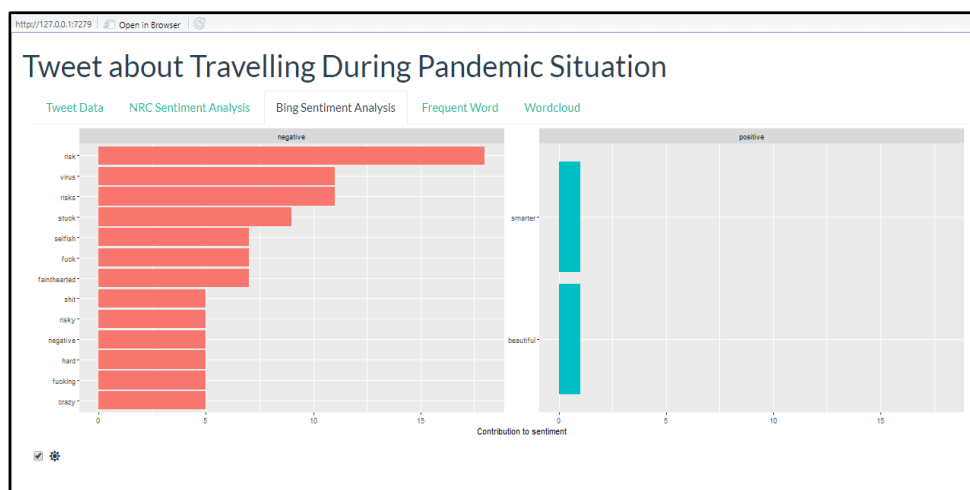
Previous **1** 2 3 4 5 ... 41 Next

☒ ☐

Gambar 3.2 Tweet Data - Light mode



Gambar 3.3 NRC Sentiment Analysis



Gambar 3.4 Bing Sentiment Analysis

4. KESIMPULAN

Berdasarkan hasil analisis, perancangan, dan pembahasan yang dilakukan, dapat disimpulkan bahwa:

- a. Media sosial twitter dapat digunakan untuk menganalisis kecenderungan sebuah opini pengguna twitter terkait topik bepergian di masa pandemi
- b. Dengan menggunakan NRC Sentiment Analysis dapat diketahui tingkat kecenderungan opini pengguna twitter dalam beberapa kategori yaitu, positif, negatif, kemarahan, antisipasi, jijik, ketakutan, kegembiraan, kesedihan, kejutan, dan kepercayaan
- c. Dengan menggunakan Bing Sentiment Analysis dapat diketahui kecenderungan opini pengguna twitter dengan mengkategorikan kata ke dalam mode biner ke dalam dua kategori yaitu positif dan negatif
- d. Berdasarkan hasil dari NRC dan Bing Sentiment Analysis kebanyakan opini yang ditulis oleh pengguna twitter mengenai topik bepergian di masa pandemi masuk ke dalam kategori negatif
- e. Berdasarkan tampilan wordcloud dan word count (Frekuensi 10 kata yang paling sering muncul) dapat dilihat bahwa kata "*people*" adalah kata yang sering muncul pada tweet tentang topik bepergian di masa pandemi

DAFTAR PUSTAKA

- [1] Hidayat, S. (2021). Analisa Sentimen Masyarakat Indonesia Terhadap Kinerja Prabowo Subianto Sebagai Anggota Kabinet Indonesia Maju di Platform Twitter Menggunakan R dengan Algoritma Naïve Bayes. *Jurnal Nasional Informatika*, 2(1), 1-10.
- [2] Permadi, V. A. (2020). Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura. *Jurnal Buana Informatika*, 11(2), 141-151. doi:10.24002/jbi.v11i2.3769
- [3] Alizah, M. D., Nugroho, A., Radiyah, U., & Gata, W. (2020). Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter. *Indonesian Journal on Software Engineering (IJSE)*, 6(2), 223–229. doi:10.31294/ijse.v6i2.8991
- [4] Fauziyyah, A. K. (2020). Analisis Sentimen Pandemi Covid-19 Pada Streaming Twitter dengan Text Mining Python. *Jurnal Ilmiah SINUS*, 18(2), 31-42. doi:10.30646/sinus.v18i2.491
- [5] Siroj, S., Arwani, I., & Ratnawati, D. (2021). Analisis Sentimen Opini Publik pada Twitter terhadap Efek Pembelajaran Daring di Universitas Brawijaya menggunakan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(7), 3131-3140.
- [6] Silge, J., & Robinson, D. (2021). *sentiment analysis with tidy data*. Welcome to Text Mining with R | Text Mining with R. Retrieved December 25, 2021, from <https://www.tidytextmining.com/sentiment.html>