

LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
ANALISIS SENTIMEN PADA MEDIA SOSIAL
TWITTER TERHADAP DONALD TRUMP



ARDLYANSYAH 123190112
DEWI ZUNUVI SETIAWATI 123190117

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

DAFTAR ISI

DAFTAR ISI	2
1. PENDAHULUAN	3
1.1 Latar Belakang	3
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1. Mengetahui bagaimana sentiment masyarakat berdasar cuitan di media sosial Twitter terhadap Donald Trump.....	3
2. Mengetahui daerah yang harus difokuskan kampanyenya.	3
2. METODE	4
2.1 Mengumpulkan Data	4
2.2 Preprocessing Data.....	4
2.3 Menentukan Metode.....	5
2.4 Klasifikasi dan Visualisasi Data	5
2.5 Pembuatan visualisasi pada Aplikasi Shiny	9
3. HASIL DAN PEMBAHASAN	13
4. KESIMPULAN	15
5. DAFTAR PUSTAKA	16

1. PENDAHULUAN

1.1 Latar Belakang

Pemilihan presiden menjadi hal yang ditunggu-tunggu, baik oleh orang-orang yang mencalonkan diri, partai politik, maupun masyarakat di suatu negara tersebut. Pemilihan ini, menjadi akhir pemerintahan sebelumnya sekaligus menjadi awal bagi pemerintahan baru yang berarti beberapa kebijakan akan diperbarui. Beberapa taktik dibuat oleh masing-masing pasangan calon untuk menarik perhatian masyarakat agar memilihnya. Namun, karena pemilihan dilakukan dengan mengumpulkan banyak pemikiran, menjadikan sulit untuk mengira-ngira pasangan calon presiden yang berkemungkinan memenangkan pemilihan.

Di Amerika sendiri, cara menentukan pemenang adalah banyaknya negara bagian yang memihak. Perhitungan kemenangan berdasarkan banyaknya negara bagian yang dimenangkan bukan dari jumlah keseluruhan masyarakat Amerika membuat tim sukses dari pasangan calon kesulitan dalam memperkirakan kemenangan. Hal ini dibutuhkan untuk membuat taktik atau langkah dalam menentukan daerah kampanye untuk menarik perhatian masyarakat setempat. Salah satu orang yang mencalonkan diri sebagai presiden adalah Donald Trump. Oleh karena beberapa hal itu, peneliti melakukan analisis sentimen masyarakat berdasar cuitan di Twitter terhadap Donald Trump sebagai langkah efektif dalam menentukan keberhasilan kampanye dengan memilih daerah yang sesuai.

1.2 Rumusan Masalah

Berdasar pada latar belakang diatas, maka dapat dirumuskan masalah sebagai berikut :

1. Bagaimana hasil *sentiment analysis* pada media sosial twitter terhadap Donald Trump?
2. Daerah mana saja yang harus difokuskan kampanyenya?

1.3 Tujuan Penelitian

Tujuan dari penelitian berdasarkan masalah yang telah dirumuskan, yaitu :

1. Mengetahui bagaimana sentiment masyarakat berdasar cuitan di media sosial Twitter terhadap Donald Trump.
2. Mengetahui daerah yang harus difokuskan kampanyenya.

2. METODE

2.1 Mengumpulkan Data

Tahap awal dalam melakukan analisis sentimen adalah menyiapkan data yang diperlukan. Pada penelitian ini, data yang digunakan pada penelitian ini adalah data sekunder yang diambil dari situs penyedia data *open source* data.world. Data dapat divalidasi keasliannya karena bersumber dari twitter. Data ini berisi 1,850,123 baris dan 6 kolom dengan bahasa yang digunakan pada tweetnya adalah bahasa Inggris.

2.2 Preprocessing Data

Preprocessing data merupakan tahapan awal setelah mendapat dataset yang diperlukan. Pada proses ini, data dilakukan pembersihan dengan mengevaluasi nilai-nilai data error atau kosong dan menghapus beberapa data yang tidak diperlukan. Hal ini dilakukan untuk meningkatkan keakuratan hasil penelitian yang akan dilakukan.

1. Menghapus tanda baca.

```
ulasan1 <- Corpus(VectorSource(ulasan))
removeURL <- function(x) gsub("http[^\s:]*", "", x)
reviewclean <- tm_map(ulasan1, removeURL)
removeNL <- function(y) gsub("\n", " ", y)
reviewclean <- tm_map(reviewclean, 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("&", "", y)
reviewclean <- tm_map(reviewclean, removeamp)
removeUN <- function(z) gsub("@\\w+", "", z)
reviewclean <- tm_map(reviewclean, removeUN)
```

Source Code 2.1 Menghapus Tanda Baca.

Karena data yang didapatkan berjumlah cukup banyak yaitu sekitar 1 jutaan, maka peneliti membatasi dengan menggunakan *sample* 10.000 data. Langkah ini dilakukan agar proses yang analisis berjalan cepat, sehingga hasil yang diperlukan dapat segera didapatkan. Dengan menggunakan **Source Code 2.1**, kita dapat menghapus tanda baca karena tanda baca tidak diperlukan dalam menganalisis sentiment.

2. Menghapus kata – kata yang tidak penting.

```
myStopwords = readLines("stopwords-en.txt")  
reviewclean <- tm_map(reviewclean, removeWords, myStopwords)
```

Source Code 2.2 Menghapus stopwords.

Source Code 2.2 digunakan untuk menghapus kata yang ada pada file stopwords-en.txt. File ini berisi kata – kata dalam bahasa inggris yang tidak penting untuk analisis sentiment. Misalnya seperti “I”, “You”, “About”, dan lain – lain.

3. Menyimpan data bersih.

```
dataframe<-data.frame(text=unlist(sapply(reviewclean, `[[`)),  
stringsAsFactors=F)  
  
View(dataframe)  
  
write.csv(dataframe, file = 'reviewclean.csv')
```

Source Code 2.3 Menyimpan File Bersih.

Setelah semua tanda baca dan kata – kata yang kurang penting pada file csv dihilangkan, kita dapat menggunakan **Source Code 2.3** untuk menyimpan file baru yang berisi data yang sudah bersih.

2.3 Menentukan Metode

Metode yang digunakan pada penelitian ini adalah metode *Naïve-bayes*. *Naïve Bayes* mengasumsikan bahwa fitur (kata) memiliki nilai yang independen pada posisi kata (Dewi et al., 2018). Cara kerja metode ini adalah dengan mengelompokkan kelas dokumen teks. Tingkat akurasi dari metode ini dirasa cukup tinggi, sehingga banyak penelitian mengenai sentiment yang menggunakan metode ini.

2.4 Klasifikasi dan Visualisasi Data

Proses yang dilakukan pada tahap ini adalah mengklasifikasikan sentimen yang kemudian hasil dari klasifikasi tersebut divisualisasikan agar pengguna lebih mudah untuk melihat.

1. Menyediakan library yang dibutuhkan.

```
library(e1071)
library(caret)
library(syuzhet)
library(ggplot2)
library(vroom)
library(here)
```

Source Code 2.4 Library yang digunakan.

Library e1071 digunakan untuk . Library caret digunakan untuk. Library syuzhet digunakan untuk. Library ggplot2 digunakan untuk menampilkan dan membuat visualisasi data berupa plot dari hasil sentiment. Library vroom digunakan untuk membuat dataset dapat digunakan hingga keseluruhan program tanpa melakukan pengulangan *import* data. Library here digunakan untuk mengarahkan letak file.

2. Melakukan visualisasi data menggunakan barplot

```
reviewData <- read.csv("reviewclean.csv", stringsAsFactors = FALSE)
review <- as.character(reviewData$text)
s <- get_nrc_sentiment(review)

review_combine <- cbind(reviewData$text, s)
par(mar=rep(3,4))
a <- barplot(colSums(s), col = rainbow(10), main='Sentiment
Analysys Donald Trump on Twitter')
brplt <- a
```

Source Code 2.5 Melakukan visualisasi dengan barplot.

Saat ingin melakukan visualisasi, hal yang pertama dilakukan adalah membaca data yang sudah dibersihkan pada tahap sebelumnya. Kemudian mengubah string pada file menjadi sebuah karakter sehingga mudah untuk mengklasifikasikannya. Kemudian, gunakan fungsi `get_nrc_sentiment` untuk mengklasifikasikan sentiment menjadi 8 bagian yaitu *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, *trust*, *negative*, dan *positive*. Setelah itu, membuat sebuah matriks yang merupakan gabungan dari text yang berada pada file `reviewData` dengan variable `s`. Mengatur parameter, dan margin. Barplot yang dihasilkan adalah jumlah kolom dari klasifikasi yang sudah dilakukan. Kemudian membuat judul dari barplot. Kemudian menampilkan barplot.

3. Menampilkan barplot dari sentimen negatif dan sentimen positif

```
#Plot 2- sentimen positif dan negatif
getsentiment <- function(x){
  #transpose
  tx<-data.frame(t(x))
  #The function rowSums computes column sums across rows for
  each level of a grouping variable.
  tx_new <- data.frame(rowSums(tx[2:900]))
  #Transformation anx cleaning
  names(tx_new)[1] <- "count"
  tx_new <- cbind("sentiment" = rownames(tx_new), tx_new)
  rownames(tx_new) <- NULL
  tx_new2<-tx_new[1:10,]
}

#Plot One - count of words associated with each sentiment
review1 <- getsentiment(s)
quickplot(sentiment, data=review1[9:10,], weight=count,
geom="bar", fill=sentiment, ylab="count")+ggtitle("Sentiment
Analysys Donald Trump on Twitter")
```

Source Code 2.7 Melakukan visualisasi positif dan negatif.

Pada bagian ini, kita hanya mengambil bagian positif dan negatif dari visualisasi sebelumnya.

4. Melakukan visualisasi data menggunakan pie chart

```
slices <- c(colSums(s))
lbls <- c("Anger", "Anticipation", "Disgust", "Fear", "Joy",
"Sadness", "surprise", "Trust", "Negative", "Positive")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct, sep = " ")
lbls <- paste(lbls,"%", sep = "")

piechart = pie(slices,labels = lbls,
col=rainbow(length(lbls)),
main="Pie Chart Sentiment Analysys Emotions")
```

Source Code 2.6 Melakukan visualisasi dengan pie chart.

Yang pertama dilakukan adalah mendefinisikan banyak potongan(bagian) pada pie chart. Banyak bagian sama dengan banyak sentiment yang dianalisis yaitu *Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Negative, dan Positive*. Selanjutnya, mengubah nilai pada masing-masing sentiment kedalam bentuk persen. Kemudian menampilkan pie chart dengan nama Pie Chart Sentiment Analysys Emotions dengan 10 warna.

5. Membuat wordcloud

```
library(tm)
library(RTextTools)
library(e1071)
library(dplyr)
library(caret)
```

Source Code 2.7 Menyediakan library.

Menyediakan library yang dibutuhkan dalam membuat wordcloud. Selain itu juga menyediakan library untuk mengklasifikasikan menggunakan *naïve bayes classifier*.

```
df<-read.csv("reviewclean.csv",stringsAsFactors = FALSE)
glimpse(df)

#Set the seed of R's random number generator, which is useful
for creating simulations or random objects that can be
reproduced.
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, 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)
```

Source Code 2.8 Melakukan klasifikasi

Membaca file review yang sudah dibersihkan sebelumnya. Kemudian menggunakan fungsi glimpse untuk melihat bentuk dari file review yang sudah dibersihkan tersebut. Kemudian melakukan cleaning file lagi untuk membersihkan file.

```
library(wordcloud)
wordcloud(corpus.clean,min.freq = 4, scale = (c(6,0.7))
,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))
```

Source Code 2.9 Menampilkan wordcloud

Menampilkan wordcloud dengan maksimal katanya berjumlah 100, serta mengatur tampilan dari wordcloud.

2.5 Pembuatan visualisasi pada Aplikasi Shiny

Hasil penelitian ini berupa aplikasi shiny.

```
library(shiny)
library(dplyr)
library(plotly)
library(here)
x = 1
sentimen <- read.csv( here("Trump_tweets_sentiment.csv"), nrow = 10000)
location <- unique(sentimen[["from_user_timezone"]])
chart = c("Emotion Analysys","Piechart","Wordcloud")
```

Source Code 3.1 Menyediakan library dan membaca file.

Menyediakan library yang dibutuhkan. Membaca file csv dan membatasi baris hingga 10000. Membuat variable location yang berisi zona waktu pengguna. Membuat variable chart yang berisi array dari visualisasi data yang ingin dilakukan.

```
ui <- fluidPage(
  title = "Sentiment Analysys",
  headerPanel("Sentiment Analysys Donald Trump on Twitter"),
  sidebarLayout(
    sidebarPanel(
      tabsetPanel(id = "tabs",
        tabPanel("Menu",
```

```

        selectInput(inputId = "location",
                    label = "location",
                    choices = location,
                    selected = location[[1]]),

        selectInput(inputId = "chart",
                    label = "Chart",
                    choices = chart,
                    selected = " ",)
      ),
    ),
  mainPanel(
    DT::dataTableOutput('tbl'),
    plotOutput('plot'),
  ))

```

Source Code 3.2 Mengatur UI.

Selanjutnya yang dilakukan adalah mengatur UI. Disini penulis menggunakan judul Sentiment Analysys, header Sentiment analysys Donald Trump on Twitter, mengatur menu untuk menampilkan visualisasi yang akan ditampilkan, menampilkan table berdasarkan menu, dan menampilkan visualisasi berdasarkan menu.

```

server <- function(input, output, session) {

  output$tbl1 = DT::renderDataTable({
    new_df <- select(sentimen, text)
    new_df2 = filter(new_df, sentimen$from_user_timezone ==
input$location)
    DT::datatable(new_df2, options = list(lengthChange =
FALSE))
  })

  output$plot = renderPlot({
    new_df <- select(sentimen, text)
    new_df2 = filter(new_df, sentimen$from_user_timezone ==
input$location)

ulasan <- new_df2$text
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(reviewclean, 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("&", "", 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-en.txt")
reviewclean <- tm_map(reviewclean, removeWords, myStopwords)

dataframe<-data.frame(text=unlist(sapply(reviewclean, `[`)),
stringsAsFactors=F)
write.csv(dataframe,file = 'nreviewclean.csv')

```

Source Code 3.3 Mengatur server.

Yang pertama dilakukan adalah menampilkan table. Table ini menampilkan data sentimen berdasarkan zona waktu yang dipilih pengguna shiny. Selanjutnya menampilkan plot. Dalam menampilkan plot, data yang awalnya utuh dipisah berdasarkan zona waktu pengguna berdasarkan input yang diberikan pengguna. Kemudian lakukan data cleaning sebelum nantinya diolah untuk ditampilkan dalam bentuk visualisasi data. Selanjutnya data yang sudah dibersihkan akan disimpan dengan nama nreviewclean.csv.

```

nreviewData <- read.csv("nreviewclean.csv",stringsAsFactors =
FALSE)
nreview <- as.character(nreviewData$text)
ns <- get_nrc_sentiment(nreview)

  if(chart[1] == input$chart){
    review_combine <- cbind(nreviewData$text ,ns)
    par(mar=rep(3,4))
    barplot(colSums(ns), col = rainbow(10), main='Sentiment
Analysys Donald Trump on Twitter')
  }

  else if(chart[2] == input$chart){
    nslices <- c(colSums(ns))
    nlbls <- c("Anger", "Anticipation", "Disgust", "Fear",
"Joy", "Sadness", "surprise", "Trust", "Negative", "Positive")
    npct <- round(nslices/sum(nslices)*100)
    nlbls <- paste(nlbls, npct, sep = " ")
    nlbls <- paste(nlbls,"%", sep = "")

    npiechart = pie(nslices,labels = nlbls,
col=rainbow(length(nlbls)),
main="Pie Chart Sentiment Analysys Emotions")
  }

  else if(chart[3] == input$chart){
    clean.for.find.word=read.csv("nreviewclean.csv")
    clean.for.find.word=clean.for.find.word$text
    find.word<- Corpus(VectorSource(clean.for.find.word))

    writeLines(strwrap(find.word[[1]]$content, 100))

    myCorpus=find.word
    tdm <- TermDocumentMatrix(myCorpus,
control = list(wordLengths = c(1, Inf)))
    term.freq <- rowSums(as.matrix(tdm))
    term.freq <- subset(term.freq, term.freq >= 10)
    df <- data.frame(term = names(term.freq), freq =
term.freq)
    df <- df[with(df, order(-freq)), ]
    n=dim(df)[1]
  }

```

```

df <- data.frame(no=1:n, df)

wordcloud(corpus.clean,min.freq = 4, scale = (c(6,0.7))
,max.words=100,random.order=F,colors=brewer.pal(8,"Dark2"))
}

},height = 600, width = 800)

}

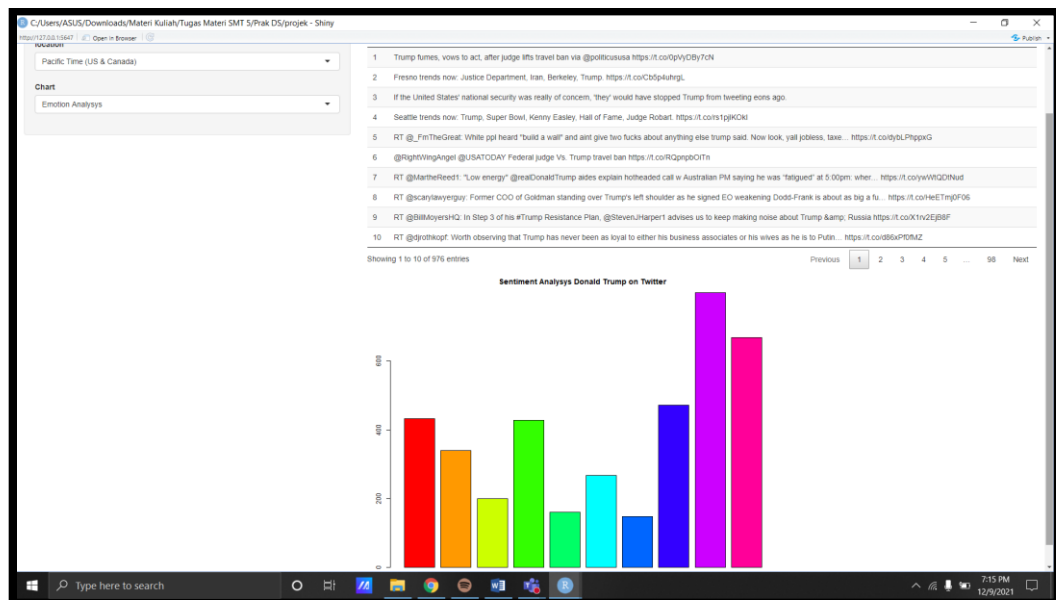
```

Source Code 3.3 Mengatur visualisasi pada shiny.

Sebelum melakukan visualisasi, kita harus membaca data yang akan digunakan. Data yang digunakan adalah data yang sudah dibersihkan sebelumnya. Selanjutnya kalimat yang ada pada data, dipecah menjadi karakter. Kemudian kata kata tersebut dikelompokkan menjadi 10. yaitu *Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Negative, dan Positive*. Selanjutnya apabila input yang diberikan sama dengan chart[1], maka yang ditampilkan adalah barplot dari kesepuluh pengelompokkan tersebut. Jika input sama dengan chart[2], maka yang ditampilkan adalah piechart dari kesepuluh pengelompokkan tersebut. Jika input sama dengan chart[3], maka yang ditampilkan adalah wordcloud dari daerah zona waktu yang ditentukan.

3. HASIL DAN PEMBAHASAN

Dari dataset yang digunakan, didapatkan hasil berupa aplikasi shiny yang dapat menampilkan hasil sentiment analisis terhadap Donald Trump dalam bentuk visualisasi grafik. Pengguna dapat melihat hasil sentimen analisis berupa grafik berdasarkan daerah yang ditentukan. Sehingga, dari hasil ini dapat mengambil langkah yang efektif dalam menentukan daerah untuk digelar kampanye. Seperti terlihat pada gambar di bawah ini



Gambar 3.1 *Histogram*

Terlihat pada aplikasi shiny bahwa negara dengan nilai negatif tinggi yaitu di US dan Canada dengan nilai sentiment negatif hingga lebih dari 600. Dengan mengubah lokasi juga bisa dilihat hasil dari penelitian ini bahwa tingkat sentiment *anger* dan negatif memiliki gap yang tidak terlalu besar dengan hasil sentimen positif. Selain itu, juga dilakukan pengelompokan secara global dengan hasil sebagai berikut.

4. KESIMPULAN

Dari penelitian ini didapatkan kesimpulan bahwa sentiment masyarakat terhadap Donald Trump dapat membantu Donald Trump menentukan pandangan masyarakat terhadapnya dan menentukan negara yang harus diberi fokus lebih saat melakukan kampanye agar tercapainya kesuksesan kampanye yang diinginkan dan mendapatkan suara lebih banyak saat pemilihan dilakukan. Tingkat sentiment positif dan negatif tidak memiliki gap yang terlalu tinggi. Namun, sentiment negative masih mewakili. Bahkan, hasil dari *wordcloud* kata-kata negatif mewakili kata-kata lainnya.

5. DAFTAR PUSTAKA

- Cahyani, Rafiqah., Indri Sudanawati Rozas, & Nita Yalina. (2020). Analisis Sentimen Pada Media Sosial Twitter. *Jurnal Ilmu Komputer dan Teknologi Informasi*, 8.
- Dewi, Y. N., Riana, D., & Mantoro, T. (2018). Improving Naïve Bayes performance in single image pap smear using weighted principal component analysis (WPCA). *3rd International Conference on Computing, Engineering, and Design, ICCED 2017*.
- Himawan, Sisferi., Amsal Pardamean, & Siti Nur Khasanah. (2020). Sentimen Analisis Publik Terhadap Joko Widodo Terhadap Wabah Covid-19 Menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, 10.