

ANALISIS SENTIMEN TWITTER PADA ISU *MENTAL HEALTH* DENGAN ALGORITMA KLASIFIKASI NAIVE BAYES

Karina Aulia¹⁾, Lia Amelia²⁾

^{1,2)} Program Studi Informatika Fakultas Teknik Universitas Siliwangi

¹⁾167006048@student.unsil.ac.id, ²⁾167006049@student.unsil.ac.id

Abstrak

Isu kesehatan mental masih menjadi satu permasalahan kesehatan yang signifikan di dunia. Upaya penurunan angka penderita penyakit mental tidak akan berlangsung sukses apabila masyarakat turut berkontribusi dalam mendiskriminasi penderita yang berada di lingkungan sekitar. Pemahaman dan stigma buruk serta kesadaran kesehatan mental yang rendah turut andil dalam upaya penyembuhan kesehatan mental. Penelitian ini bertujuan menerapkan algoritma naïve bayes dalam proses klasifikasi data tweets terkait isu kesehatan mental. bertujuan mencari tahu bagaimana stigma yang berkembang di masyarakat mengenai isu kesehatan mental. Pencarian opini dilakukan dengan mencari data *tweets* dari Twitter menggunakan Twitterscraper. Data diklasifikasikan dengan algoritma Naive Bayes untuk mengelompokkan sentimen dan emosi. Hasil eksperimen menunjukkan bahwa sentimen positif yang mendominasi.

Kata kunci: Analisis Sentimen, Crawling, Kesehatan Mental

Abstract

The issue of mental health is still a significant health problem in the world. Reporting from the publication of the Indonesian Ministry of Health, according to WHO data (2016), there were about 35 million people affected by depression, 60 million people affected by bipolar disorder, 21 million affected by schizophrenia, and 47.5 million affected by dementia. Unfortunately, according to a research journal, efforts to reduce the number of people with mental illness will not be successful if the community contributes to discriminating sufferers who are among them. Poor understanding and stigma and low mental health awareness contribute to efforts to cure mental health. For this reason, this study aims to find out how the stigma develops in society regarding mental health issues. An opinion search is done by searching the tweets data from Twitter using Twitterscraper with the Python programming language. The data is then classified with the Naive Bayes algorithm to classify sentiments and emotions. The classification process uses Rstudio with the R programming language. The experimental results show positive sentiment that dominates.

Keywords— Sentimen Analysis, Crawling, Mental Health.

I. PENDAHULUAN

Isu kesehatan mental masih menjadi satu permasalahan kesehatan yang signifikan di dunia. Dilansir dari publikasi Kemenkes RI, menurut data WHO (2016), terdapat sekitar 35 juta orang terkena depresi, 60 juta orang terkena bipolar, 21 juta terkena skizofrenia, serta 47,5 juta terkena demensia [1]. Data lain dari riset kesehatan dasar pada tahun 2007 di Indonesia terdapat sekitar 1 juta orang yang mengalami gangguan jiwa berat dan 19 juta orang dengan gangguan jiwa ringan hingga sedang, dengan jumlah yang terus meningkat secara signifikan [2].

Berbagai upaya rehabilitasi dan diskusi keilmuan ditempuh organisasi terkait untuk menurunkan jumlah penderita penyakit mental ini. Hal tersebut tidak akan

berlangsung sukses apabila masyarakat turut berkontribusi dalam mendiskriminasi penderita yang berada di antara mereka [3]. Pemahaman dan stigma yang buruk serta kesadaran kesehatan mental yang rendah turut andil dalam upaya penyembuhan kesehatan mental. Secara umum, stigma penyakit mental dapat dikonseptualisasikan sebagai stigma publik dan stigma diri. Stigma publik dipahami sebagai stigma dari masyarakat umum yang mewakili stereotip, prasangka, dan diskriminasi yang berkaitan dengan penyakit mental [4].

Salah satu cara untuk mendapatkan opini mengenai stigma yang berkembang di masyarakat umum terhadap isu kesehatan mental adalah melalui media sosial. Twitter Inc. sebagai perusahaan jejaring sosial daring terkemuka di dunia mengungkapkan ada 126

juta pengguna aktif setiap harinya selama kuartal empat (Q4) pada 2018 lalu [5]. Sehingga pada penelitian ini akan dilakukan analisis sentimen dengan menganalisis *tweets* dari pengguna Twitter mengenai *mental health*. Untuk menganalisis sentimen yang dikemukakan oleh pengguna Twitter mengenai isu *mental health* akan dilakukan proses klasifikasi dengan algoritma Naive Bayes. Algoritma Naive Bayes dipilih karena memiliki performa akurasi yang lebih baik [6].

II. METODOLOGI PENELITIAN

Eksperimen ini memiliki beberapa tahapan metodologi.

Metode yang dilakukan antara lain:

a. Studi kepustakaan

Metode pengumpulan data dengan metode kepustakaan dilakukan dengan pengumpulan jurnal, literatur, *paper*, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan metode *rule-based*.

b. Proses *crawling* data

Data yang diperoleh merupakan sumber data yang diambil secara langsung dari Twitter menggunakan bahasa Python dan *library* *Twitterscraper* dengan keyword pencarian "*mental health*".

```
twitterscraper "mental health" -  
-limit 100000 - -
```

Data yang didapatkan dari proses *crawling* ini adalah sebanyak 82.715 *tweets* berbahasa Inggris dan format *.json*. Data ini kemudian diimport ke dalam *database* MongoDB.

c. Praproses data sentimen

Data mentah yang diperoleh dilakukan pemrosesan untuk mengubahnya menjadi data yang siap digunakan. Tahap praproses data ini, dilakukan secara manual dan menggunakan *tools* Rstudio. Tahapan yang dilakukan yaitu:

1) Penghapusan *tweet* berulang dengan perintah berikut:

```
#Case Folding  
> twitclean <-  
  tm_map(twitclean, tolower)  
>twitclean <-  
  tm_map(twitclean,  
  stripwhitespace)
```

2) Proses *tokenizing* yaitu penghapusan URL,

tanda baca, dan simbol-simbol yang tidak diperlukan dari sebuah *tweets*.

```
#--- TOKENIZING -----  
-----#  
#build Corpus  
uncleanedTweets<-  
Corpus(VectorSource(unique_twe  
et))
```

```
removeURL <- function(x)  
gsub("http[^\s:]*", "",  
x)  
twitclean<-  
tm_map(uncleanedTweets,  
removeURL)
```

```
removeNL <- function(y)  
gsub("\n", " ", y)  
twitclean <-  
tm_map(twitclean, removeNL)
```

```
replacecomma <- function(y)  
gsub(",", "", y)  
twitclean <-  
tm_map(twitclean,  
replacecomma)
```

```
removeRT <- function(y)  
gsub("RT ", "", y)  
twitclean <-  
tm_map(twitclean, removeRT)
```

```
removetitik2 <- function(y)  
gsub(":", "", y)  
twitclean <-  
tm_map(twitclean,  
removetitik2)
```

```
removetitikkoma <-  
function(y) gsub(";", " ", y)  
twitclean <-  
tm_map(twitclean,  
removetitikkoma)
```

```
removetitik3 <- function(y)  
gsub("p...", "", y)  
twitclean <-  
tm_map(twitclean,  
removetitik3)
```

```
removeamp <- function(y)  
gsub("&", "", y)  
twitclean <-
```

```
tm_map(twitclean, removeamp)

removeUN <- function(z)
gsub("@\\w+", "", z)
twitclean <-
tm_map(twitclean, removeUN)

remove.all<-function(xy)
gsub("[^[:alpha:][:space:]]*"
, "", xy)
twitclean <-
tm_map(twitclean,remove.all)

#Menghapus titik koma,
menjadi non kapital
twitclean <-
tm_map(twitclean,
removePunctuation)
```

- 3) Proses *case folding*: menyeragamkan huruf menjadi kecil.

```
#Case Folding
> twitclean <-
tm_map(twitclean, tolower)
>twitclean <-
tm_map(twitclean,
stripwhitespace)
```

- 4) *Stopword removal* yaitu penghapusan kata imbuhan yang ada dalam sebuah *tweet*.

```
#Stopwords Removal/Filtering
twitclean <-
tm_map(twitclean,
removewords,
c(stopwords("english"), 'just',
'da', 'that', 'live',
'your', 'isnt', 'etc',
'didnt'))
```

Keempat proses tersebut menjadikan data *tweets* yang siap dipakai sebanyak 77.458 *tweets*. Data ini kembali diinputkan ke dalam *database* MongoDB dalam format csv untuk kemudian dilakukan proses klasifikasi.

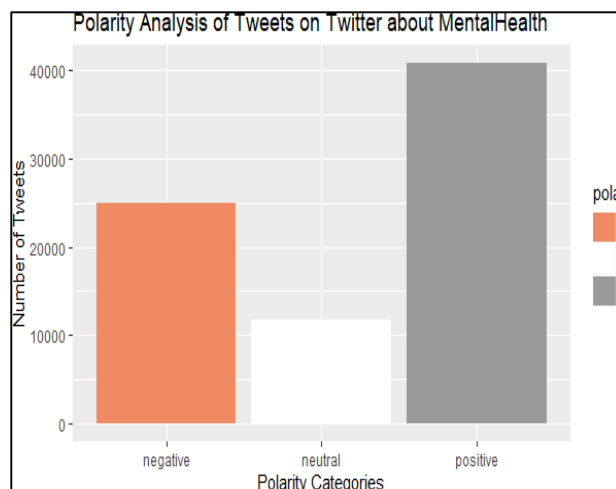
III.HASIL DAN PEMBAHASAN

Eksperimen dilakukan dengan melakukan proses klasifikasi menggunakan algoritma Naive Bayes. Data yang digunakan adalah data *tweets* yang sebelumnya sudah diinputkan ke dalam *database*.

- a. Klasifikasi polaritas

Pada eksperimen ini dilakukan proses

pengklasifikasian polaritas dimana *tweets* dikelompokkan berdasarkan sentimen positif, netral, dan negatif. Hasil ini muncul berdasarkan skor perhitungan dari algoritma yang digunakan. Gambar 1 menampilkan hasil skor 20 *tweets* teratas.



Gambar 1. Skor 20 *tweets* teratas

Dalam Tabel 1, ditampilkan contoh masing-masing *tweets* yang mengandung sentimen positif, netral, dan negatif.

Tabel 1. Tweets dengan klasifikasi sentimen

Tweet	Klasifikasi
<i>Looking forward enjoying mental health today</i>	positif
<i>Many chat windows detrimental ones mental health oh well one live go</i>	Netral
<i>Still tired think will blow work mental health day im overdue</i>	negatif

Sedangkan hasil klasifikasi keseluruhan *tweets* divisualisasikan dalam gambar 2.

```
> #measure polarity (pos, neg, neutral):
> TweetsClassPol = classify_polarity(hasil_cleaning, algorithm = "bayes")
There were 50 or more warnings (use warnings() to see the first 50)
> #lihat hasilnya:
> head(TweetsClassPol, 20)
```

	POS	NEG	POS/NEG	BEST_FIT
[1,]	"1.03127774142571"	"9.47547003995745"	"0.108836578774127"	"negative"
[2,]	"25.670707456461"	"9.47547003995745"	"2.70917509613869"	"positive"
[3,]	"1.03127774142571"	"8.78232285939751"	"0.117426534862834"	"negative"
[4,]	"1.03127774142571"	"8.78232285939751"	"0.117426534862834"	"negative"
[5,]	"8.78232285939751"	"17.8123396772424"	"0.493047124551418"	"negative"
[6,]	"1.03127774142571"	"9.47547003995745"	"0.108836578774127"	"negative"
[7,]	"26.3638546370209"	"26.8423564950873"	"0.982173627037777"	"negative"
[8,]	"1.03127774142571"	"9.47547003995745"	"0.108836578774127"	"negative"
[9,]	"1.03127774142571"	"0.445453222112551"	"2.31512017476245"	"positive"
[10,]	"9.47547003995745"	"8.78232285939751"	"1.07892526745566"	"neutral"
[11,]	"1.03127774142571"	"8.78232285939751"	"0.117426534862834"	"negative"
[12,]	"8.78232285939751"	"0.445453222112551"	"19.7154772340574"	"positive"
[13,]	"17.2265151579293"	"0.445453222112551"	"38.6718836070664"	"positive"
[14,]	"1.03127774142571"	"17.1191924966825"	"0.0602410272345239"	"negative"
[15,]	"1.03127774142571"	"17.8123396772424"	"0.0578968153601572"	"negative"
[16,]	"1.03127774142571"	"0.445453222112551"	"2.31512017476245"	"positive"
[17,]	"16.5333679773693"	"9.47547003995745"	"1.74485992859976"	"neutral"
[18,]	"32.0354582133129"	"0.445453222112551"	"71.9165484119422"	"positive"
[19,]	"8.78232285939751"	"0.445453222112551"	"19.7154772340574"	"positive"
[20,]	"9.47547003995745"	"0.445453222112551"	"21.2715265477174"	"positive"

Gambar 2. Hasil klasifikasi keseluruhan tweets

Hasil klasifikasi menunjukkan tweets dengan sentimen positif mendominasi dengan perolehan data lebih dari 40.000 tweets, disusul dengan sentimen negatif lebih dari 25.000 tweets dan netral lebih dari 10.000 tweets.

b. Klasifikasi emosi

Pada tahapan ini, tweet diberi skor ke dalam 6 klasifikasi emosi yaitu *anger* (marah), *disgust* (jijik), *fear* (takut), *joy* (kegembiraan), *sadness* (kesedihan), dan *surprise* (mengejutkan). Skor tertinggi akan ditampilkan pada kolom *best fit* (paling cocok). Hasil ini muncul berdasarkan skor perhitungan dari algoritma yang digunakan. Gambar 3 dan gambar 4 menampilkan hasil skor dari 20 tweets teratas.

	ANGER	DISGUST	FEAR
[1,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[2,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[3,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[4,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[5,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[6,]	"1.46871776464786"	"3.09234031207392"	"7.34083555412328"
[7,]	"1.46871776464786"	"7.34083555412328"	"2.0678359955953"
[8,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[9,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[10,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[11,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[12,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[13,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[14,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[15,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[16,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[17,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[18,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[19,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"
[20,]	"1.46871776464786"	"3.09234031207392"	"2.0678359955953"

Gambar 3. Skor 20 tweets dengan emosi marah, jijik, dan takut

	JOY	SADNESS	SURPRISE	BEST_FIT
[1,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[2,]	"7.34083555412328"	"1.7277074477352"	"2.78695866252273"	"joy"
[3,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[4,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[5,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[6,]	"1.02547755260094"	"1.7277074477352"	"7.34083555412327"	"fear"
[7,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	"disgust"
[8,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[9,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[10,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[11,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[12,]	"7.34083555412328"	"1.7277074477352"	"2.78695866252273"	"joy"
[13,]	"7.34083555412328"	"1.7277074477352"	"2.78695866252273"	"joy"
[14,]	"1.02547755260094"	"7.34083555412328"	"2.78695866252273"	"sadness"
[15,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[16,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA
[17,]	"7.34083555412328"	"1.7277074477352"	"2.78695866252273"	"joy"
[18,]	"13.6561935556456"	"1.7277074477352"	"2.78695866252273"	"joy"
[19,]	"7.34083555412328"	"1.7277074477352"	"2.78695866252273"	"joy"
[20,]	"1.02547755260094"	"1.7277074477352"	"2.78695866252273"	NA

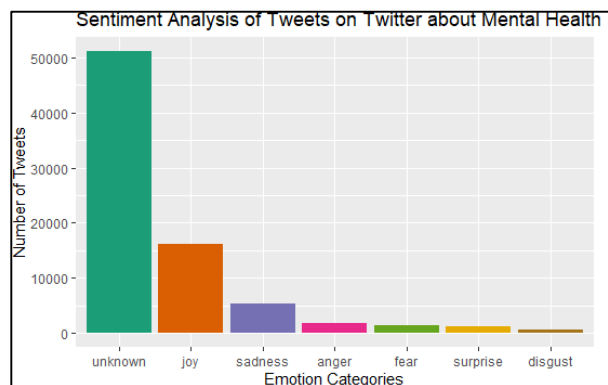
Gambar 4. Skor 20 tweets dengan emosi bahagia, sedih, terkejut, dan hasil klasifikasi paling cocok.

Dalam Tabel 2, ditampilkan data tweets dengan hasil klasifikasi emosinya. Selain keenam klasifikasi emosi di atas, terdapat juga tweet dengan klasifikasi emosi yang tidak diketahui karena skor keenam emosi yang memiliki range tidak jauh.

Tabel 2. Tweets dengan klasifikasi emosi

Tweet	Klasifikasi
<i>Looking forward enjoying mental health today</i>	tidak diketahui
<i>Many chat windows detrimental ones mental health oh well one live go</i>	bahagia
<i>Still tired think will blow work mental health day im overdue</i>	Jijik
<i>Number prescription given children depression mental health problems</i>	sedih
<i>Yesterday awful talking mental health day</i>	takut

Sedangkan hasil klasifikasi keseluruhan tweets divisualisasikan dalam gambar 5.



Gambar 5. Tweets dengan klasifikasi emosi

- indonesia/. [Diakses 15 Oktober 2019].
- [3] P. Corrigan, "Mental Health Stigma as Social Attribution: Implications," *Clinical Psychology Science and Practice*, vol. 7, no. 1, pp. 2-4, 2006.
- [4] Y. Pratomo, "Tekno Kompas," *Kompas*, 09 02 2019. [Online]. Available: <https://tekno.kompas.com/read/2019/02/09/11340027/untuk-pertama-kali-twitter-ungkap-jumlah-pengguna-harian>. [Accessed 15 Oktober 2019].
- [5] L. Crumb, T. M. Mingo and A. Crowe, "Get over it and move on": The impact of mental illness stigma in rural, low-income United States populations," *Mental Health & Prevention*, vol. 13, pp. 143-148, 2019.
- [6] M. R. Maarif, "Perbandingan Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Judul Artikel," *Jurnal Informatika Sunan Kalijaga*, vol. 1, no.2, p. 1, 2016.