

Classification of Sentiments on Twitter Opinions with The Keyword Sinovac Using Naive Bayes

Klasifikasi Sentimen pada Opini Twitter dengan Kata Kunci Sinovac Menggunakan Naive Bayes

Bagus Muhammad Akbar¹, Ahmad Taufiq Akbar², Rochmat Husaini³

^{1,2,3} Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

¹bagusmuhammadakbar@upnyk.ac.id, ²ahmadtaufiq.akbar@upnyk.ac.id,

³husaini@upnyk.ac.id

Abstract

Keywords: Data Mining, Sentiment Analysis, Naive Bayes, Valence Shifter, Lexicon, Vaccine, Covid-19, Sinovac

Purpose:Applying the sentiment analysis method using the Naive Bayes method on Twitter opinions regarding public sentiment towards the covid-19 vaccine, especially the Sinovac type.

Design/Methodology/Approach: Valence Shifter-Lexicon based and Data mining with Bayes algorithm to determine sentiment category about Sinovac vaccine

Findings/result: Sentiment analysis using the Naive Bayes method resulted in 1433 (71.65%) positive sentiment, 403 (20.15%) negative sentiment, and 164 (8.2%) neutral sentiment. Meanwhile, sentiment analysis using the Valence Shifter-lexicon based method resulted in 903 (45.15%) positive sentiment, 437 (21.85%) negative sentiment, and 660 (33%). Based on sentiment analysis with these 2 methods, the Naive Bayes method is more significant in classifying sentiment. In addition, the results of this study also indicate that the emergence of the Sinovac vaccine gives a positive impression among the public.

Originality/Value/State of the Arts:

The methods used in this study are 2 methods, namely Valence Shifter-lexicon based and bayes to determine public sentiment about the Sinovac vaccine. So that clarifying the results of the sentiment analysis tend to be positive even though the labeling is not done manually.

Kata kunci: Data Mining, Analisis Sentimen, Naive Bayes, Valence Shifter, Lexicon, Vaksin, Covid-19, Sinovac

Abstrak

Tujuan :Menerapkan metode analisis sentimen menggunakan metode Naïve Bayes pada opini twitter mengenai sentimen masyarakat terhadap vaksin covid-19, khususnya jenis sinovac.

Metode: Valence Shifter-Lexicon based dan Data mining dengan algoritma Naïve Bayes untuk menentukan kategori sentimen tentang vaksin sinovac

Hasil : Analisis sentimen menggunakan metode Naïve Bayes menghasilkan 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) untuk sentimen netral. Sedangkan analisis sentimen menggunakan metode Valence Shifter-lexicon based menghasilkan 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebesar 660 (33%).. Berdasarkan analisis sentimen dengan 2 metode tersebut, metode Naive bayes lebih signifikan dalam mengklasifikasikan sentimen. Disamping itu, Hasil penelitian ini juga mengisyaratkan bahwa kemunculan vaksin sinovac memberikan kesan positif di kalangan masyarakat.

State of the Art:

Metode yang digunakan dalam penelitian ini adalah 2 metode yakni Valence Shifter-lexicon based dan Naïve Bayes untuk menentukan sentimen masyarakat tentang vaksin sinovac. Sehingga lebih mengklarifikasi hasil analisis sentimen cenderung positif meskipun tidak dilakukan pelabelan secara manual.

1. Pendahuluan

Hingga di awal tahun 2021 Indonesia masih terkena dampak dari penyebaran pandemi COVID-19 yang telah melanda di beberapa negara sejak awal tahun 2020 [1][2]. Dalam masa itu, berbagai vaksin telah mulai digunakan oleh berbagai negara di dunia termasuk Indonesia untuk penanganan pandemi covid-19[2][3]. Salah satu vaksin yang marak diperbincangkan adalah sinovac yang pertama kali diperkenalkan pemerintah indonesia untuk mengantisipasi dampak pandemi covid -19 ini [4]. Sebelum vaksin covid-19 diberlakukan di Indonesia, Pemerintah Indonesia melalui Kementerian Kesehatan bersama WHO-UNICEF telah melakukan survei terhadap 112,888 responden dari berbagai provinsi, hasilnya sebanyak 65% bersedia menerima vaksin yang akan diberlakukan pemerintah, 8% tidak setuju untuk divaksin, dan 27 % masih mempertimbangkan kelayakan vaksin [1]. Peredaran vaksin covid-19 juga menuai respon pro dan kontra di berbagai media sosial, salah satunya twitter. Respon tersebut bermunculan karena tidak sedikit yang mengkhawatirkan dampak dari vaksinasi dan validitas dari uji kelayakan klinis beberapa vaksin covid-19 yang beredar [2].

Berdasarkan fenomena tersebut, akan lebih baik jika perkembangan opini di media sosial diteliti dan dianalisis sehingga diketahui trend sentimen masyarakat terhadap vaksin covid-19. Penelitian mengenai analisis sentimen dari kumpulan opini pada media twitter telah dilakukan pada sejumlah penelitian sebelumnya. Penelitian mengenai analisis sentimen terhadap transportasi umum jakarta menggunakan metode Naive Bayes menghasilkan akurasi pengujian sebesar 95,88% dengan 10 data uji dan 90 data latih [5]. Metode Naive Bayes dan Support Vector Machine telah digunakan dalam penelitian untuk analisis sentimen dari data twitter yang dikelompokkan menurut fitur unigram dan bigram [6]. Metode Naive Bayes digunakan dalam analisis sentimen untuk deteksi emosi dengan mempertimbangkan beberapa fitur yakni fitur linguistik, fitur ortografik dan Fitur n-gram, namun hanya fitur n-gram yang sangat signifikan berpengaruh pada peningkatan akurasi, sedangkan fitur linguistik dan ortografik hampir tidak meningkatkan akurasi karena sedikitnya kemunculan dari fitur tersebut pada seluruh data [7]. Diluar metode Naive Bayes, analisis sentimen twitter juga dapat dilakukan melalui analisis frekuensi n-gram dan penilaian polaritas sehingga diketahui nilai sentimen dominan dari setiap opini [8]. Analisis sentimen terhadap vaksin covid-19 juga telah dilakukan menggunakan data twitter sejak januari 2021 sebanyak 6000 opini berbahasa Indonesia, setelah melalui pelabelan dan uji klasifikasi Naive Bayes, diperoleh hasil sentimen negatif sebesar 56%, sentimen positif 39 %, dan sentimen netral 1%. Ini menunjukkan bahwa trend sentimen tergantung pada lokasi geografis dan periode data pengambilan tersebut [1]. Analisis sentimen vaksin covid-19 dilakukan dalam penelitian yang mengambil data twitter sejak Oktober hingga November 2020, diperoleh trend bahwa sentimen positif lebih unggul 6% dibandingkan sentimen negatif [2]. Penelitian tentang analisis sentimen pada vaksin sinovac di Filipina dari data tweet selama periode Maret 2021, menggunakan ekstraksi fitur n-gram kemudian direpresentasikan dalam word vector dengan TF-IDF dan diklasifikasikan menggunakan Naive Bayes menghasilkan dominasi sentimen positif beserta akurasi yang sangat signifikan [3]. Penelitian lain [9], telah menguji kinerja analisis sentimen dari 5 macam R *library* yang menunjukkan bahwa *library sentimentr* memiliki akurasi tertinggi, melebihi 4 *library* lain seperti *RSentiment*, *syuzhet*, *meanr*, dan *SentimentAnalysis*.

Dengan demikian, penelitian ini akan menggunakan metode *valence shifter-lexicon based* yang terdapat pada R *Package* yakni *library sentimentr* dan metode Naive Bayes yang terdapat pada *library sentiment*, dengan tujuan dapat menganalisis sentimen masyarakat terkait vaksin khususnya sinovac dalam opini (tweet) berbahasa Inggris. Meskipun tanpa melalui pelabelan manual, kedua metode diharapkan dapat mengklarifikasi hasil trend sentimen terhadap vaksin sinovac sehingga masyarakat lebih bijaksana dalam menyikapi informasi terkait himbauan vaksin.

2. Metode/Perancangan

Metode penelitian ini terdiri atas beberapa tahap yakni akuisisi data berupa pengambilan data tweet (opini) yaitu melalui *crawling* data dari Twitter. Dilanjutkan dengan tahap pre-processing, yang meliputi *case folding* (pengubahan ukuran huruf), *normalisasi data* tweet, penghapusan *stopword*, *tokenizing* serta *stemming*. Tahap berikutnya adalah analisis sentimen. Analisis sentimen dilakukan dengan perhitungan menurut *valence shifter* dan *lexicon*, yaitu perhitungan skor polaritas dari bagian-bagian kalimat dalam setiap opini, sehingga polaritas pada setiap opini dapat ditentukan sebagai sentimen positif, negatif atau netral[9]. Kemudian, analisis sentimen

dengan metode kedua berupa klasifikasi Naïve Bayes yang menghitung probabilitas kemunculan setiap term atau kata indikator dalam setiap opini yang menunjuk sentimen positif, neutral atau negatif sehingga probabilitas yang tertinggi menjadi label sentimen pada opini yang dihitung[3]. Setelah dilakukan pengujian terhadap data tweet maka akan muncul hasil berupa proporsi sentimen positif, netral dan negatif[11]. Setelah itu, dilakukan visualisasi hasil analisis sentimen dari kedua metode beserta evaluasi akhir.

2.1. Pengumpulan Data

Metode penelitian ini diawali dengan proses akuisisi data melalui *crawling* data tweet dengan kata kunci “sinovac” yang muncul terbaru sejak 1 Februari 2021. Pengumpulan data dilakukan dengan memanfaatkan API twitter, yang menggunakan 5 komponen otentikasi yaitu consumer key, cosumer secret, access token, dan access token secret [10]. Data twitter yang dikumpulkan berdasarkan kata kunci “sinovac” dengan *setting* bahasa Inggris, dan diambil mulai tanggal 1 Februari 2021 dalam kategori tweet terbaru dan tweet terpopuler sebanyak 2000 data tweet. Proses pengambilan dilakukan dengan mendownload data tweet terkait vaksin covid-19 yaitu “sinovac” menggunakan *library* `Twitter` di aplikasi RStudio. Data disimpan dalam file .csv dan akan masuk pada tahap preprocessing data.

2.2. Preprocessing Data

Data yang telah disimpan dalam file .csv akan masuk ke proses preprocessing data. Tahap ini meliputi pembersihan data pada komponen-komponen tertentu pada setiap komentar tweet yaitu username, URL (Uniform Resource Locator), dan karakter khusus seperti hashtag (#), retweet (RT), tanda baca (.,?! dll), lainnya karakter (\$,%*, dll). Karena nama pengguna, URL, karakter khusus, tanda baca, dan karakter lainnya tidak berdampak signifikan terhadap nilai sentimen, komponen tipikal harus dibersihkan dengan menghapus atribut dalam kutipan di Twitter. Komponen nama pengguna diidentifikasi dengan kemunculan karakter “@”. Selain username, karakter “@” juga bisa digunakan untuk menelepon tempat-tempat seperti @CinemaXXI. Namun, nama tempat tidak berpengaruh pada analisis sentimen, sehingga nama tempat juga harus dihilangkan. Komponen URL dikenali melalui ekspresi reguler (seperti http, www) [12].

Preprocessing data ini dilakukan dalam beberapa tahap antara lain, *case folding* (pengubahan ukuran huruf), *normalisasi data* tweet, penghapusan *stopword*, *tokenizing* serta *stemming*. Tahapan preprocessing adalah [13] :

- a. *Case folding* atau pengubahan karakter huruf (hanya sebatas ‘a’ hingga ‘z’) kedalam huruf kecil (non kapital), selain huruf akan dihapus karena dianggap sebagai delimiter.
- b. Normalisasi, yakni menghapus komponen delimiter, tag html dan emoji, karena komponen tersebut tidak diperlukan dalam analisis sentimen,
- c. Penghapusan *stopword*, ini dilakukan dengan pemeriksaan kata ganti, kata depan atau kata-kata yang terdapat dalam pustaka *stopword* berbahasa Inggris, sehingga, jika ditemukan kata-kata yang sama dalam pustaka tersebut, maka akan dihapus .
- d. *Stemming* yakni menghilangkan imbuhan pada setiap kata sehingga diubah kedalam bentuk kata dasar.
- e. Tokenisasi, berupa pemotongan kalimat berdasarkan kata-kata yang menyusunnya dan karakter-karakter pemisah kata.

2.3. Analisis Sentimen

Analisis Sentimen atau opinion mining adalah studi komputasi untuk mengidentifikasi dan mengungkapkan pendapat, sentimen, evaluasi, sikap, emosi, subjektivitas, penilaian, atau pandangan yang terkandung dalam sebuah teks [14]. Analisis sentimen juga merupakan metode untuk mengekstrak data opini, memahami dan memproses data tekstual secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini [15]. Analisis sentimen dalam penelitian ini menggunakan 2 metode, pertama metode *valence shifter* dan *lexicon*, kemudian yang kedua adalah metode Naive Bayes.

2.3.1. Valence Shifter dan Lexicon

Metode ini memperhitungkan nilai bobot yang terkandung pada setiap kata positif atau negatif dari kalimat dalam suatu opini (tweet), kemudian menjumlahkan bobot kata-kata tersebut berdasarkan pustaka lexicon dengan memperhitungkan pula kemunculan *Valence-Shifter* yang meliputi, *negator* (misalkan: "never", "none"), *amplifier* (misalkan: "very"), *de-amplifier* (misalkan: "few", "little bit"), *adversative* (misalkan: "but", "instead") dan *contrasting conjunctions* ((misalkan: "however"), sehingga diketahui nilai selisih antara polaritas positif dan negatif, jika selisih sama dengan nol maka nilai sentimen akhir dari opini tersebut diklasifikasikan sebagai sentimen netral [13][16]. Metode *Valence Shifter* dan *lexicon* ini dilakukan menggunakan fungsi `sentimen_by()` pada *library sentimentr* dari *package Rstudio*.

2.3.2. Naive Bayes

Metode Naive Bayes sebagai bentuk aplikasi Teori Bayes, merupakan metode klasifikasi yang dapat memprediksi kelas berdasarkan probabilitas kelas dan probabilitas setiap objek terhadap kelas [17]. Naive Bayes dapat mengambil keputusan berdasarkan data yang diperoleh sebelumnya. Meskipun asumsi independensi antar kata (term) dalam dokumen tidak dapat dipenuhi sepenuhnya, kinerja Naive Bayes dalam klasifikasi relatif sangat baik [8]. Naive Bayes memiliki keunggulan dari segi kemudahan, kecepatan klasifikasi, dan akurasi yang cukup tinggi. Naive Bayes juga dapat digunakan sebagai alat pengambilan keputusan, salah satunya tentang klasifikasi sentimen [1][3].

Algoritma Naive Bayes yang digunakan dalam penelitian ini terdapat dalam fungsi `classify_polarity()` dalam *sentiment library* yang akan mengklasifikasikan opini kedalam 3 kelas, yaitu sentimen positif, netral, dan sentimen negatif. Hasil klasifikasi diperoleh dengan menghitung probabilitas dari sebuah tweet dengan asumsi berada pada salah satu dari dua kelas yang ada. Setelah nilai probabilitas dihitung, maka dihitung juga rasio antara probabilitas terhadap kelas positif dengan probabilitas terhadap kelas negatif, dan dari hasil rasio ini akan diperoleh nilai untuk mengklasifikasikan tweet pada salah satu kelas yang ada. Klasifikasi ke dalam kelas netral diperoleh jika hasil perhitungan rasio di atas, nilainya sama dengan 1[18].

3. Hasil dan Pembahasan

Setelah mengumpulkan data menggunakan *library twitter* melalui aplikasi RStudio, maka diperoleh 2000 tweet berbahasa Inggris. Jenis tweet yang diambil adalah tweet terbaru dan tweet terpopuler yang terkait dengan kata kunci pencarian "sinovac". Pada Tabel 1 merupakan fungsi dalam bahasa R untuk mengambil data tweet melalui proses *crawling*.

Tabel 1. Source code untuk tweet data mining

```
#mining data twitter
miningtweets <- searchTwitter('sinovac',
                             lang="en", n=2000,since='2021-02-01',resultType = "mixed");
miningtweets_text <- sapply(miningtweets,function(x) x$getText());
#Tempat penyimpanan file csv
pathOutput = "E:\\Unduhan\\coba\\wi\\"
write.csv(miningtweets_text, paste(pathOutput,'dataTwitterv2.csv',sep = ""))
```

Selanjutnya, untuk setiap opini, semua ukuran huruf akan diubah menjadi huruf kecil. Hal ini sangat penting dilakukan karena kata-kata yang terdiri dari huruf yang sama akan dibaca oleh sistem sebagai kata yang berbeda jika huruf-huruf tersebut memiliki ukuran huruf yang berbeda. Misalnya seperti kata 'Happy' dan 'happy'. Jika hal ini dibiarkan akan berdampak pada bertambahnya waktu komputer untuk memproses data.

Seperti terlihat pada Tabel 1, fungsi *x* dipanggil untuk melakukan konversi ke huruf kecil, sehingga tidak ada kata yang sama tetapi dianggap berbeda karena huruf kapital pada kata tersebut. Selanjutnya akan muncul data hasil *crawling* seperti pada Tabel 2 berupa kumpulan data opini (tweet) dengan id record 1 sampai 2000. Setiap opini masih mengandung URL, hashtag, berbagai tanda baca sehingga harus dihilangkan agar kata-kata yang terkumpul pada setiap opini menjadi kata dasar sehingga diperoleh lebih banyak kata yang mengerucut pada nilai skor kategori sentimen atau polaritas tertentu.

Tabel 2 . Crawled data

```
"52","@marqdmartian Can I take another brand other than Sinovac after 6 months?"
"53","RT @pharmaniaga_bhd: Kelantan will receive 40,000 doses of the Sinovac vaccine on Thursday,
said Kelantan Health Department director Datuk..."
"54","@R04ch You mentioned a well-known YouTuber escaping Dodge to get her own. Actress
""Chompoo"" Hargate appears to be h... https://t.co/CvyDvB5AtO"
"55","RT @TruthAbtChina: There is a vaccine shortage in China and vaccinations are first come, first
served.

The sign above the door says, ""COVI..."
"56","expert reaction to announcement of results of phase III of Sinovac vaccine in China | Science Media
Centre https://t.co/5T1OBq5Htv"
"57","RT @khalidkarim: Indonesia have thus far received 84 mil doses of Sinovac.Yesterday , they
received another 8 mil doses.
```

Penghapusan tanda baca dan URL yang tidak penting dilakukan dengan beberapa fungsi di *library* *twitter* seperti terlihat pada Tabel 3.

Tabel 3. Fungsi Data Cleaning pada tweet

```
miningtweets_text = read.csv(paste(pathOutput,'dataTwitterv2.csv',sep = ""))
miningtweets_text = gsub("https.*", "", miningtweets_text$x)
miningtweets_text = gsub("&", " ", miningtweets_text)
miningtweets_text = gsub("\n", " ", miningtweets_text)
miningtweets_text = sub("^\\s*<U\\+|\\w+>\\s*", "", miningtweets_text)
# remove retweet entities
miningtweets_text = gsub("(RT|via)((?:\\b\\W*@[\\w+])+", " ", miningtweets_text)
# remove at people
```

```

miningtweets_text = gsub("@\\w+", " ", miningtweets_text)
# remove punctuation
miningtweets_text = gsub("[^\\x01-\\x7F]", " ", miningtweets_text)
miningtweets_text = gsub("[[:punct:]]", " ", miningtweets_text)
# remove numbers
miningtweets_text = gsub("[[:digit:]]", " ", miningtweets_text)
# remove html links
miningtweets_text = gsub("http\\w+", " ", miningtweets_text)
# remove unnecessary spaces
miningtweets_text = gsub("[ \\t]{2,}", " ", miningtweets_text)
miningtweets_text = gsub("^\\s+|\\s+$", " ", miningtweets_text)
miningtweets_text = gsub("note", " ", miningtweets_text)
    
```

Penghapusan *stopwords* juga dilakukan dengan menggunakan fungsi dari *tm* library sebagaimana pada Tabel 4 untuk menghapus stopwords dalam bahasa Inggris. *Stopword* harus dihapuskan karena pada penelitian sebelumnya menunjukkan bahwa pengujian sentimen analisis dari koleksi opini yang tidak mengandung *stopword* memiliki tingkat akurasi yang lebih tinggi daripada analisis sentimen pada kumpulan opini yang masih mengandung *stopword* [19].

Tabel 4. Fungsi Stopword Remover

```

library(tm)
all_tweet<-datalengkap
all_tweets <- as.data.frame(all_tweet)
komen <- all_tweets$text
komenc <- Corpus(VectorSource(komen))
twitclean <- komenc
twitclean <- tm_map(twitclean, removeWords, stopwords("english"))
twitclean_new <- tm_map(twitclean, removePunctuation)
    
```

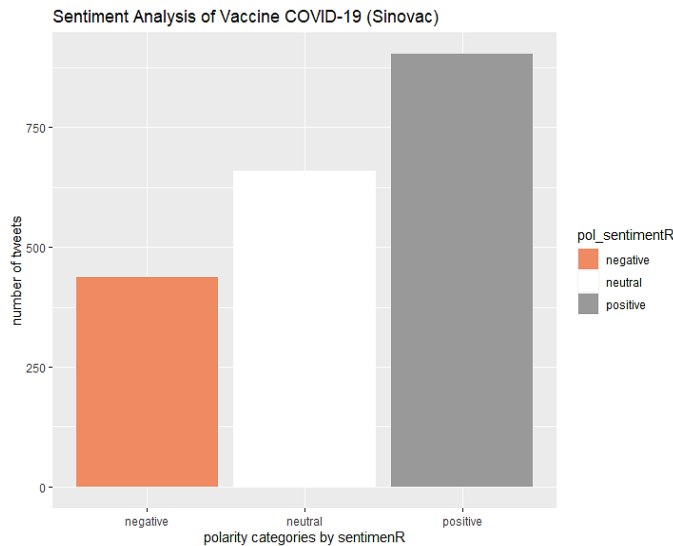
Hasil dari seluruh proses pembersihan data akan mengubah setiap kata menjadi token seperti yang terlihat pada Tabel 5, sehingga setiap opini bersih dari notasi dan imbuhan yang tidak diperlukan pada analisis sentimen .

Tabel 5 . Hasil dari proses *preprocessing*

Tweet Sebelum Preprocessing	Tweet setelah Pre-processing
"52","@marqdmartian Can I take another brand other than Sinovac after 6 months?"	"52"," can i take another brand other than sinovac after months "
"53","RT @pharmaniaga_bhd: Kelantan will receive 40,000 doses of the Sinovac vaccine on Thursday, said Kelantan Health Department director Datuk..."	"53"," kelantan will receive doses of the sinovac vaccine on thursday said kelantan health department director datuk "
"54","@R04ch You mentioned a well-known YouTuber escaping Dodge to get her own. Actress ""Chompoo"" Hargate appears to be h... https://t.co/CvyDvB5AtO "	"54"," you mentioned a well known youtuber escaping dodge to get her own actress chompoo hargate appears to be h "
"55","RT @TruthAbtChina: There is a vaccine shortage in China and vaccinations are first come, first served. The sign above the door says, ""COVI..."	"55"," there is a vaccine shortage in china and vaccinations are first come first served the sign above the door says covi "

"56","expert reaction to announcement of results of phase III of Sinovac vaccine in China Science Media Centre https://t.co/5T1OBq5Htv"	"56","expert reaction to announcement of results of phase iii of sinovac vaccine in china science media centre "
"57","RT @khalidkarim: Indonesia have thus far received 84 mil doses of Sinovac. Yesterday , they received another 8 mil doses.	"57"," indonesia have thus far received mil doses of sinovac yesterday they received another mil doses why they can get "

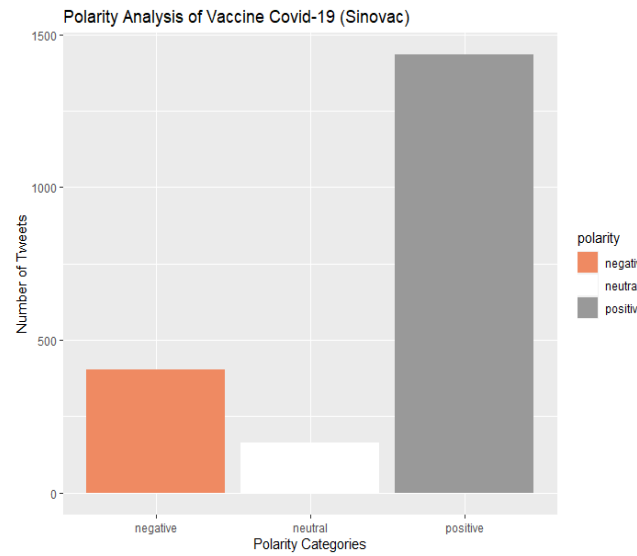
Data tweet yang telah bersih kemudian diklasifikasikan menggunakan metode *Valence Shifter* dan *Lexicon* melalui fungsi `sentiment_by()` dari *library sentimentr* sehingga diperoleh kategori sentimen (polaritas) yakni negatif, netral dan positif sebagaimana tampak pada Gambar 6.



Gambar 6. Hasil Analisis Sentimen dari Fungsi `Sentiment_by()`

Fungsi `sentiment_by()` dari *library sentimentr* mengklasifikasikan sentimen melalui metode *Valence Shifter* dan *Lexicon* dengan menghitung skor polaritas rata-rata seluruh teks atau kelompok tertentu teks, dan akan memberikan nilai NA jika text hanya 1 kalimat [20]. Hasil klasifikasi sentimen menggunakan fungsi `sentiment_by` ditunjukkan pada Gambar 6 yakni sebanyak 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebesar 660 (33%). Pada Gambar 6 terlihat bahwa sebagian besar sentimen masyarakat terhadap kata kunci “sinovac” adalah positif dan netral.

Selain itu, penelitian ini juga menggunakan *library sentiment* dengan memanggil fungsi `classify_polarity()` yang menggunakan metode *Naïve Bayes*. Hasil klasifikasi sentimen dari *library sentiment* ditunjukkan pada Gambar 7.



Gambar 7 . Hasil Klasifikasi Menggunakan fungsi `Classify_polarity`

Sebagaimana pada Gambar 7., diperoleh hasil sentimen analisis sebanyak 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) untuk sentimen netral. Berdasarkan penelitian sebelumnya[9], *library sentimentr* (Metode *Valence Shifter* dan *Lexicon*) memiliki keunggulan dalam menentukan polaritas sehingga dapat meminimalkan munculnya polaritas netral. Namun jika dibandingkan dengan hasil dari Gambar 7 terlihat bahwa jumlah polaritas netral dari hasil *library sentiment* (metode *Naïve Bayes*) lebih sedikit dan polaritas positif lebih besar. Sedangkan hasil sentimen negatif pada kedua metode tersebut relatif setara. Jika merujuk pada penelitian tersebut[9], dalam penelitian kali ini *library sentiment* lebih akurat, karena hasil sentimen negatif relatif setara pada kedua metode namun *library sentiment* lebih mampu membedakan antara polaritas negatif dan positif. Ini juga menguatkan kemungkinan bahwa sebagian besar sentimen netral dari hasil analisis sentimen oleh *library sentimentr* (metode *valence Shifter* dan *Lexicon*) memiliki polaritas positif. Sebagaimana pada Gambar 6 dan Gambar 7, terlihat bahwa *library sentiment* (Metode *Naïve Bayes*) lebih signifikan dalam membedakan sentimen polaritas positif dan negatif, terutama pada kata kunci “sinovac”.

Sentimen netral yang lebih banyak pada hasil metode *Valence shifter–lexicon based* ini sangat dimungkinkan karena fungsi penilaian pada metode tersebut akan menggabungkan bobot yang dikaitkan dengan kemunculan *Valence-shifter*, termasuk *negator*, *amplifier* (penguat), *de-amplifier* (*downtoner*), dan konjungsi adversatif. Pengaruh kemunculan *Valence Shifter* ini dapat memodifikasi tanda dan intensitas polaritas dalam sebuah teks. Misalnya, *negator* atau konjungsi adversatif dapat membalikkan atau mengabaikan seluruh orientasi semantik kalimat(arah sentimennya). Jika frekuensi kemunculan *Valence-shifter* sangat tinggi, pencocokan kata berdasarkan *lexicon* akan sulit memodelkan orientasi semantik secara akurat. Hal ini berbeda dengan kemampuan *machine learning* yakni *Naïve Bayes* yang mendeteksi sentimen berdasarkan probabilitas term yang mengarah pada sentimen positif atau sentimen negatif.

4. Kesimpulan dan Saran

Berdasarkan hasil analisis sentimen yang telah dilakukan, kedua metode baik *Naive Bayes* maupun *Valence Shifter-lexicon based* menunjukkan bahwa masyarakat cenderung merespon positif terhadap vaksin Sinovac berdasarkan analisis terhadap komentar tweet pada kata kunci sinovac selama pandemi covid-19 sejak Februari 2021.

Analisis sentimen menggunakan *library sentiment* (Metode Naïve Bayes) berhasil mendeteksi polaritas positif dengan persentase tertinggi, dan polaritas netral dengan persentase terendah dibandingkan dengan *library sentimentr* (Metode *Valence Shifter-lexicon based*). Berdasarkan kedua metode tersebut, analisis sentimen juga menunjukkan sentimen positif terhadap kata kunci sinovac, karena pada Februari 2021 pemberitaan vaksin sinovac membawa tren positif dalam penanganan covid-19. Untuk penelitian selanjutnya sebaiknya analisis sentimen terhadap data tweet juga dikaitkan dengan klasifikasi emosi dari tweet, sehingga dapat diketahui pengaruh emosi pada sentimen opini.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia yang telah memberikan dana untuk penelitian ini.

DAFTAR PUSTAKA

- [1] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.
- [2] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175>.
- [3] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- [4] "Dimulainya Vaksinasi Covid-19 di Indonesia Halaman all - Kompas.com." <https://nasional.kompas.com/read/2021/01/14/06572221/dimulainya-vaksinasi-covid-19-di-indonesia?page=all> (accessed Oct. 02, 2021).
- [5] D. I. Sari, Y. F. Wati, and Widiastuti, "Analisis Sentimen Dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum Mrt Jakarta Menggunakan Naïve Bayes Classifier," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 64–75, 2020, doi: 10.35760/ik.2020.v25i1.2427.
- [6] I. E. Syahputra, T. Tulus, and S. Efendi, "Indonesian Text Dataset for Determining Sentiment Classification Using Machine Learning Approach," *J. Informatics Telecommun. Eng.*, vol. 3, no. 2, pp. 192–201, 2020, doi: 10.31289/jite.v3i2.3153.
- [7] I. Fera Fanesya, Randy Cahya Wihandika, "Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, p. 3, 2019.
- [8] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.
- [9] M. Misuraca, A. Forciniti, G. Scepi, and M. Spano, "Sentiment Analysis for Education with R: packages, methods and practical applications," no. 2008, 2020, [Online]. Available: <http://arxiv.org/abs/2005.12840>.
- [10] K. Sigit, A. P. Dewi, G. Windu, N. N. Nurmalasari, T. Muhamad, and N. Kadinar, "Comparison of Classification Methods on Sentiment Analysis of Political Figure Electability Based on Public Comments on Online News Media Sites," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 662, no. 4, 2019, doi: 10.1088/1757-899X/662/4/042003.
- [11] A. Mustofa Hidayat and M. Syafrullah, "Algoritma Naïve Bayes Dalam Analisis Sentimen Untuk Klasifikasi Pada Layanan Internet PT.XYZ," *J. Telemat. MKOM*, vol. 9, no. 2, pp. 91–95, 2017, [Online]. Available: <http://journal.budiluhur.ac.id/index.php/telematika/article/view/532>.
- [12] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi:

- 10.35314/isi.v3i1.335.
- [13] G. N. Aulia and E. Patriya, "Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019," *J. Ilm. Inform. Komput.*, vol. 24, no. 2, pp. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.
- [14] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*, 2010, pp. 627–666.
- [15] A. Sari, F. V., & Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 2, pp. 681–686, Nov. 2019, Accessed: Sep. 22, 2021. [Online]. Available: <https://jurnal.umk.ac.id/index.php/simet/article/view/3487>.
- [16] S. Balbi, M. Misuraca, and G. Scepti, "Combining different evaluation systems on social media for measuring user satisfaction," *Inf. Process. Manag.*, vol. 54, no. 4, pp. 674–685, 2018, doi: 10.1016/j.ipm.2018.04.009.
- [17] N. Saputra, "(Sentiment Analisis With Lexicon Preprocessing)," *Din. Inform.*, vol. 7, no. 1, pp. 45–57, 2019.
- [18] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," pp. 105–112, 2003, doi: 10.3115/1119355.1119369.
- [19] A. F. Hidayatullah, "Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016.
- [20] M. E. Wu, S. Y. Chang, C. J. Lu, and H. M. Sun, "A communication-efficient private matching scheme in Client-Server model," *Inf. Sci. (Ny)*, vol. 275, pp. 348–359, 2014, doi: 10.1016/j.ins.2014.01.016.