



ANALISIS SENTIMEN DAN EMOSI VAKSIN SINOVAC PADA TWITTER MENGGUNAKAN NAÏVE BAYES DAN VALENCE SHIFTER

Bagus Muhammad Akbar¹, Ahmad Taufiq Akbar², Rochmat Husaini³

^{1,2,3}Jurusan Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta
Yogyakarta, D.I. Yogyakarta, Indonesia

bagusmuhammadakbar@upnyk.ac.id, ahmadtaufiq.akbar@upnyk.ac.id, husaini@upnyk.ac.id

Abstract

The Sinovac vaccine is among the Covid-19 news in the world in early 2021. That information has led to public responses between the pros and cons. Through Twitter media, the public responds to the issue of the Sinovac; therefore, their opinions on Twitter can analyze to count the percentage of sentiment and emotion towards the Sinovac. This analysis hopes to provide a wise and objective reference, although the pros and cons information is still emerging. This study uses Rstudio for sentiment analysis through Twitter opinion classification using Naïve Bayes and the Valence Shifter Lexicon method to analyze emotions, also using the Naïve Bayes. The Data is 2000 English-language Twitter comments limited to the latest and most popular tweet based on the Sinovac keyword since February 1, 2021, from all Twitter users worldwide. The results showed that Naïve Bayes recognized 1433 (71.65%) positive sentiments, 403 (20.15%) negative sentiments, and 164 (8.2%) neutral sentiments. Meanwhile, Valence Shifter Lexicon recognized 903 (45.15%) positive sentiment, 437 (21.85%) negative sentiment, and 660 (33%) neutral sentiments. The Naïve Bayes also succeeded in recognizing emotions with the highest number 1727 (86.35%) mixed emotions and 141 (7.05%) joy emotion.

Keywords: Sentiment, Naïve Bayes, Valence, Covid-19, Sinovac

Abstrak

Vaksin Sinovac termasuk dalam berita tentang Covid-19 di dunia sejak awal tahun 2021. Informasi jenis vaksin ini memunculkan berbagai tanggapan publik yang tidak luput dari pro dan kontra. Melalui media sosial Twitter, masyarakat dunia juga menanggapi persoalan vaksin Sinovac sehingga kumpulan opini pada Twitter tersebut dapat dianalisis untuk mengetahui persentase sentimen dan emosi terhadap vaksin Sinovac. Dengan analisis sentimen dan emosi ini diharapkan dapat memberikan referensi yang bijak dan objektif meskipun informasi pro dan kontra masih beredar. Penelitian ini menggunakan aplikasi Rstudio untuk analisis sentimen melalui klasifikasi opini Twitter dengan metode Naïve Bayes dan Valence Shifter Lexicon. Untuk analisis emosi menggunakan klasifikasi dengan metode Naïve Bayes. Data penelitian merupakan kumpulan komentar Twitter berbahasa Inggris sebatas 2000 data terbaru dan terpopuler berdasarkan *keyword* Sinovac sejak 1 Februari 2021 dari seluruh pengguna Twitter di dunia. Hasil penelitian menunjukkan Naïve Bayes mengenali sebanyak 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) sentimen netral. Sedangkan Valence Shifter Lexicon mengenali sebanyak 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebesar 660 (33%). Metode Naïve Bayes juga berhasil mengenali emosi dengan jumlah terbanyak yakni emosi campuran (*mixed*) 1727(86,35%) dan emosi *joy* (gembira) sebanyak 141 (7,05%).

Kata kunci: Sentimen, Naïve Bayes, Valence, Covid-19, Sinovac

1. PENDAHULUAN

Informasi penanganan pandemi Covid-19 pada awal tahun 2021 melalui vaksinasi telah menyebar luas di berbagai negara [1]. Vaksin Sinovac merupakan jajaran vaksin yang pertama diperkenalkan dan menjadi pembicaraan oleh kalangan publik di dunia [2]. Hal ini juga menimbulkan berbagai tanggapan yang menuai pro dan kontra terkait kehandalan dan profil tentang vaksin Sinovac ini [3][4].

Tanggapan masyarakat di berbagai media sosial untuk saling menyampaikan aspirasi tentang vaksin Sinovac juga bermunculan salah satunya pada media sosial Twitter [3][2]. Jumlah pengguna Twitter meningkat sangat pesat terhitung pada tahun 2019 sebesar 166 juta pengguna yang terdaftar [5]. Perilaku masyarakat dunia cukup terbiasa mengunggah komentar (*tweet*) terhadap suatu fenomena yang sedang hangat menjadikan daya tarik tersendiri [6].

Oleh karena itu, untuk dapat mengetahui sentimen dan emosi pengguna Twitter terkait isu pada vaksin Sinovac, maka perlu dilakukan analisis sentimen dan emosi dari kumpulan opini atau *tweet*.

Sentiment Analysis (Analisis Sentimen) atau Analisis Sentimen atau sering disebut juga dengan *opinion mining* (penambangan opini) adalah proses ekstraksi opini dan mengolah melalui perhitungan untuk identifikasi ekspresi opini, arah sentimen, emosi, subjektivitas, dan penilaian yang terdapat pada teks opini [7][8]. Emosi merupakan salah satu aspek dalam kecerdasan manusia dan berperan penting untuk mengambil keputusan dalam kehidupan sehari-hari. Dalam interaksi manusia dengan komputer banyak mengabaikan hubungan emosional yang terjadi padahal hal tersebut sangat penting dalam komunikasi antar manusia pada kehidupan kita sehari-hari. Terdapat enam kategori emosi dasar pada manusia yakni, *happy* (senang), *sad* (sedih), *fear* (takut), *anger* (marah), *surprise* (terkejut), dan *disgust* (jijik) [9].

Dengan analisis sentimen dan emosi maka kumpulan data opini pengguna Twitter mengenai vaksin Sinovac akan diketahui seberapa banyak yang mengandung sentimen positif, negatif atau netral, dan juga kategori emosinya. Hasil dari analisis tersebut dapat memberikan penilaian terhadap suatu isu sehingga akan memudahkan dalam membuat keputusan yang bijak dan objektif [6].

Dalam menganalisis emosi dan sentimen pada teks atau kumpulan kutipan opini dapat melalui metode *data mining* berupa mesin pembelajaran maupun metode Lexicon [10]. Algoritma Naïve Bayes adalah salah satu mesin pembelajaran (*Machine Learning*) yang berdasarkan teori Bayes [11][12]. Algoritma ini dan telah digunakan sebagai metode dalam melakukan klasifikasi terhadap data kumpulan opini media sosial termasuk Twitter sehingga dapat mengetahui persentase setiap jenis sentimen yang terkandung didalamnya [3][5][6][13].

Naïve Bayes dapat melakukan klasifikasi melalui prediksi probabilitas setiap kelas [15][16]. Naïve Bayes dapat memberikan kesimpulan berdasarkan klasifikasi data-data latih yang telah tersimpan sebelumnya. Walaupun sifat independensi antar kata (*term*) atau parameter dalam dokumen tidak terpenuhi secara sempurna, namun performa Naïve Bayes untuk mengklasifikasikan relatif handal bahkan terbilang unggul dari segi kepraktisan, kecepatan, dan akurasi yang tinggi [17].

Selain metode Naïve Bayes, *Support Vector Machine* (SVM) juga telah digunakan dalam klasifikasi sentimen data *tweet* mengenai vaksin Covid-19, namun kecepatan klasifikasinya masih dibawah Naïve Bayes [6]. Analisis sentimen terhadap vaksin Sinovac dari data *tweet* berbahasa Indonesia dengan metode Naïve Bayes telah menghasilkan persentase sentimen positif yang lebih besar daripada sentimen negatif [18]. Fluktuasi jumlah sentimen positif ataupun negatif juga tergantung pada waktu

pengambilan opini *tweet* dan kondisi geografis dimana data opini berasal [1][6]. Penelitian mengenai analisis sentimen vaksin Covid-19 oleh Akbar et al. 2021 menggunakan metode Naïve Bayes dan Valence Shifter Lexicon telah menunjukkan hasil sentimen positif yang tertinggi dari kumpulan data *tweet* yang diperoleh sejak Februari tahun 2021. Namun dalam penelitian tersebut tidak melibatkan klasifikasi emosi pada data *tweet* sehingga belum diketahui bagaimana hubungan setiap kategori emosi terhadap jenis sentimennya [3]. Penelitian lain oleh Misuraca et al. 2020 telah menunjukkan bahwa metode Valence Shifter Lexicon yakni pada *library* sentiment memiliki akurasi yang tertinggi dalam mengklasifikasikan sentimen dari data teks dibandingkan 4 *library* lainnya yakni *Library* RSentiment, *syuzhet*, *meanr*, dan *SentimentAnalysis* [19].

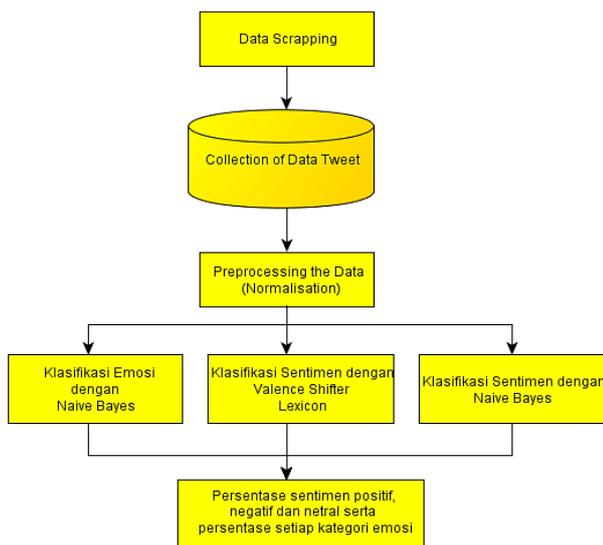
Adapun batasan dalam penelitian ini antara lain, data yang digunakan diambil dari Twitter yang dibatasi sejumlah 2000 *tweet* dari pengguna Twitter di seluruh dunia, berbahasa Inggris, yang terbaru dan terpopuler sejak 1 Februari 2021, berdasarkan kata kunci "sinovac". Data *tweet* yang diambil tidak dibatasi dengan rentang waktu karena API Twitter akan memilih sebanyak 2000 data *tweet* yang terbaru dan terpopuler sebagaimana *setting scrapping* yang digunakan pada penelitian ini.

Penelitian dilakukan dengan *tools* Rstudio menggunakan metode Naïve Bayes dan metode Valence Shifter Lexicon. Valence Shifter Lexicon adalah metode untuk menghitung polaritas sentimen yang terkandung pada teks dengan memperhitungkan bobot kemunculan kata-kata yang berasosiasi dengan setiap kategori sentimen berdasarkan Lexicon, dan juga memperhitungkan bobot dari kemunculan kata-kata dari *Valence Shifter* [10]. Lexicon adalah pustaka kata (*term*) yang berasosiasi dengan kategori sentimen ataupun kategori emosi [19]. Kedua metode analisis sentimen baik Naïve Bayes dan Valence Shifter Lexicon digunakan agar dapat dibandingkan sekaligus dapat menegaskan perolehan sentimen dominan dari masyarakat global mengenai vaksin Sinovac apakah negatif atau positif. Dengan demikian, akan dapat membantu memberikan pandangan objektif untuk memutuskan secara bijak walaupun isu pro dan kontra masih menyelimuti banyak jenis vaksin tidak terkecuali Sinovac. Disamping itu, kumpulan data *tweet* dalam penelitian ini juga diklasifikasikan menurut kategori emosinya agar dapat mengungkap emosi publik terkait informasi vaksin Sinovac, sehingga juga diketahui hubungan antara kategori sentimen dan kategori emosinya.

2. METODE PENELITIAN

Metode penelitian ini melalui beberapa tahap sebagaimana Gambar 1. yakni pengumpulan data Twitter secara *scrapping* atau *crawling* berdasarkan kata kunci "sinovac" melalui aplikasi Rstudio. Dilanjutkan dengan *Preprocessing* data *tweet* melalui normalisasi. Setelah data menjadi bersih, kemudian data diklasifikasi menggunakan Naïve Bayes dengan memanggil fungsi

`classify_polarity()` dari library `sentiment` dan juga diklasifikasikan menggunakan `Valence Shifter Lexicon` melalui fungsi `sentiment_by()` dari library `sentimentr` untuk mengetahui persentase kategori sentimen dari seluruh data.



Gambar 1. Alur Proses Penelitian

Disamping itu, data juga diklasifikasi berdasarkan emosi menggunakan `Naive Bayes` melalui fungsi `classify_emotion()` dari library `sentiment` sehingga akan diperoleh persentase setiap kategori emosi dari seluruh data *tweet*.

2.1 Metode Pengumpulan Data, Instrumen Penelitian, dan Metode Pengujian

Pengunduhan data menggunakan aplikasi Rstudio dengan memanfaatkan API Twitter melalui beberapa kunci autentikasi yakni *consumer key*, *consumer secret key*, *bearer_token key*, *access token key*, dan *access secret key* sehingga dapat diizinkan oleh Twitter untuk mengunduh data-data komentar Twitter sebagai data penelitian ini. Data Twitter yang diteliti diperoleh melalui *crawling* atau *scrapping* dari kumpulan *tweet* seluruh pengguna Twitter di dunia, berdasarkan kata kunci “sinovac”, berbahasa Inggris, dan terbatas sejak 1 Februari 2021 pada *tweet* terbaru, terpopuler dengan batas sebanyak 2000 data *tweet*.

Setelah data diunduh, sejalan dengan alur penelitian pada Gambar 1, data dinormalisasi agar menjadi bersih dan mudah untuk diklasifikasikan. Kemudian dilakukan klasifikasi pada data dengan metode `Naive Bayes` dan `Valence Shifter Lexicon`. Pengujian dilakukan dengan membandingkan hasil klasifikasi kedua metode tersebut. Penelitian ini tidak melakukan pelabelan secara manual pada data agar tidak memakan waktu yang cukup lama sehingga untuk memaksimalkan validitas dari hasil analisis sentimen digunakanlah pembandingan yaitu metode `Valence Shifter Lexicon` yang didukung dari library `sentimentr` pada Rstudio [3]. Library tersebut telah terbukti handal

berdasarkan pengujian akurasi klasifikasi sentimen [19]. Pengujian berikutnya melalui klasifikasi emosi pada data sehingga pemetaan kategori emosi menurut kategori sentimennya dapat dianalisis.

2.2 Tahapan Penelitian

1. Data Crawling (Pengumpulan Data)

Pengunduhan data dilakukan menggunakan platform Rstudio dengan Library (Twitter) untuk menjalankan proses *crawling* atau *scrapping* terhadap kumpulan *tweet* (opini) yang berkaitan covid-19 dengan keyword “sinovac”. Data yang diunduh berupa kumpulan komentar (*tweet*) dari para pengguna Twitter yang kemudian data tersebut disimpan dalam file berekstensi .csv untuk diproses pada tahap normalisasi (*preprocessing*).

2. Preprocessing (Normalisasi)

Preprocessing merupakan tahapan awal yang sangat penting sebelum dilakukan proses *data mining*. *Preprocessing* akan mengeliminasi sejumlah permasalahan diantaranya *missing value*, *data redundant*, *outliers*, ataupun format data yang tidak sesuai dengan sistem dan berpotensi dapat mengurangi performa proses *data mining*. Setelah dilakukan *preprocessing* diharapkan data telah direpresentasikan dalam kondisi yang ideal untuk diproses. Dalam analisis sentimen yang memerlukan klasifikasi data berupa dokumen teks perlu didahului dengan *preprocessing* berupa normalisasi data yang meliputi beberapa tahapan yakni *case folding*, *filtering (eliminate punctuation)*, *stopword removal*, *stemming*, *tokenization* dan sebagainya [15] [16].

Data hasil *crawling* yang telah disimpan sebagai file .csv kemudian diproses pada tahap *preprocessing* dengan menghilangkan unsur-unsur (komponen) khusus yang tidak berguna pada setiap opini (*tweet*), sehingga dicapai representasi data *tweet* lebih mudah untuk dianalisis sentimennya oleh algoritma *mining*. Ada sejumlah komponen khusus yang sering terdapat pada data opini (*tweet*) diantaranya username yang ditandai kemunculan karakter ‘@’, kemudian komponen URL (*Uniform Resource Locator*) yang ditandai dengan `http://www.`, dan karakter khas meliputi *hashtag* (#), *retweet*(RT), tanda baca (.,?! dll), karakter lainnya (\$,%*,#, dll). Komponen-komponen khas tersebut tidak berdampak signifikan terhadap klasifikasi sentimen, maka sejumlah atribut komponen tersebut harus dibersihkan dari setiap kutipan di Twitter yang telah terkumpul. Karakter “@” juga digunakan untuk merujuk ke suatu tempat seperti “@Senayan”. Tetapi nama tempat tersebut tidak berdampak signifikan pada analisis sentimen atau justru menjadi *noise* yang hanya membebani memori proses, sehingga juga perlu dihilangkan [14]. Penghapusan sejumlah komponen yang tidak berdampak tersebut melalui proses normalisasi data dengan tahap sebagai berikut:

- a. Pemeriksaan pada data opini (*tweet*) untuk mendeteksi keberadaan *username*, URL, karakter khas, tanda baca, dan karakter lainnya. Jika komponen-komponen tersebut ditemukan maka akan dihapus.
- b. Melakukan *Case Folding*. Tahap ini akan mengubah semua karakter menjadi *lower case* atau huruf kecil. Proses *Case Folding* dilakukan dengan mendeteksi ukuran setiap karakter dari awal hingga akhir setiap data opini. Jika karakter terdeteksi sebagai huruf kapital (*uppercase*), maka akan diubah menjadi huruf kecil (*lowercase*).
- c. Penghapusan *stopword* berupa daftar kata yang tidak berdampak pada analisis sentimen
- d. Penghapusan spasi yang berlebihan pada komentar *tweet*
- e. Penghapusan data *tweet* yang kosong atau NA

3. Klasifikasi Sentimen

Setelah dilakukan normalisasi, kemudian data *tweet* disimpan sebagai file.csv. Dengan hasil pembersihan ini maka akan memudahkan Naïve Bayes maupun Valence Shifter Lexicon dalam mengenali sentimen setiap kalimat *tweet*. Properti yang digunakan untuk klasifikasi sentimen yakni *library sentimentr* dengan memanggil fungsi `sentiment_by()` yang menggunakan Valence Shifter Lexicon sehingga diperoleh nilai analisis sentimen dari setiap data opini, jika nol maka tergolong dalam sentimen netral, jika > 0 maka sentimen positif dan sebaliknya adalah sentimen negatif.

Adapun properti lain dari Rstudio yang juga digunakan untuk klasifikasi sentimen pada riset ini adalah *library sentiment* melalui fungsi `classify_polarity()` yang mengimplementasikan Naïve Bayes sehingga seluruh data *tweet* akan diketahui jenis polaritasnya (kategori sentimennya). Klasifikasi sentimen atau polaritas dalam bentuk positif dan negatif menggunakan `classify_polarity()` dalam *library sentiment* sebelumnya telah diuji menggunakan metode Naïve Bayes menurut Leksikon subjektivitas Janyce Wiebe dan untuk mempelajari pola ekstraksi dari ekspresi subjektif [20].

Kedua *library* tersebut digunakan dalam penelitian karena untuk pelabelan jenis polaritas (kategori sentimen) terhadap sebanyak 2000 data *tweet* akan memakan waktu yang tidak sedikit jika dilakukan secara manual. Sehingga hasil dari fungsi `sentiment_by()` milik *library sentimentr* dapat digunakan sebagai pembandingan karena memiliki akurasi yang telah terbukti mengungguli akurasi klasifikasi dari fungsi-fungsi pada sejumlah *library* yang lain yakni *library Rsentiment*, *meanr*, *syuzhet*, dan *SentimenAnalysis*, menurut penelitian sebelumnya [19].

4. Klasifikasi Emosi

Fungsi `classify_emotion()` pada *library sentiment* juga digunakan dalam penelitian ini, untuk melakukan klasifikasi emosi seperti *anger*, *disgust*, *fear*, *joy*, *sadness*, dan *surprise*. Fungsi tersebut menggunakan metode Naïve Bayes dan telah teruji pada kumpulan *dataset* sekitar 1500 kata oleh Carlo Strapparava dan Alessandro Valitutti [21]. Ketika probabilitas total dari kata-kata dalam suatu data opini yang mengarah pada setiap emosi, tidak memiliki selisih yang signifikan, maka kategori emosi yang dilabelkan pada data opini tersebut adalah *mixed*.

Fungsi `get_nrc_sentiment()` dari *library syuzhet* juga digunakan dalam penelitian ini untuk klasifikasi dan menampilkan distribusi dari jenis-jenis emosi yang tidak dapat ditampilkan oleh *library sentiment* pada fungsi `classify_emotion`. Fungsi `get_nrc_sentiment` akan memanggil kamus sentimen NRC untuk akumulasi keberadaan 7 kategori emosi berbeda dan valensinya yang sesuai dalam file teks. Kamus NRC adalah daftar istilah bahasa Inggris yang terkait dengan polaritas positif atau negatif beserta salah satu dari delapan emosi (*anger*, *fear*, *anticipation*, *trust*, *surprise*, *sadness*, *happy*, dan *disgust*). Asumsi yang mendasari adalah bahwa, terlepas dari beberapa perbedaan budaya, mayoritas kategori emosional stabil di semua bahasa [19]. Leksikon tersedia lebih banyak dari seratus bahasa, menerjemahkan istilah asli melalui *Google Translate*.

Setiap kata dalam komentar opini (*tweet*) yang telah dibersihkan akan berasosiasi pada kategori emosi tertentu. Sehingga ekstraksi fitur akan lebih mudah dilakukan pada *tweet* yang setiap kata-katanya telah menjadi token (kata tunggal). Proses ekstraksi fitur emosi pada fungsi `get_nrc_sentiment` dari *library syuzhet* menggunakan metode Leksikon (pustaka kata yang berasosiasi dengan sentimen positif atau negatif) yang berasosiasi juga dengan kamus (pustaka) emosi. Pustaka yang dipanggil pada fungsi ini adalah kamus sentimen NRC atau *emolex* [9]. Satu kata dalam *tweet* dapat berasosiasi pada beberapa jenis emosi. Contoh pada kata 'virus' berasosiasi pada tipe emosi 'marah' (*anger*), 'jijik' (*disgust*), 'takut' (*fear*), 'sedih' (*sadness*), dan 'terkejut' (*surprise*). Misalkan *Do dataset* opini, dan *K* adalah kalimat opini pada *dataset Do*, maka dapat kita tuliskan $Do = [K_1, K_2, K_3, \dots, K_n]$ dengan *n* adalah banyaknya opini. Jika *W* adalah kata yang terdapat pada setiap kalimat opini *K* maka berlaku $K = [W_1, W_2, \dots, W_m]$ dengan *m* adalah banyaknya kata dalam kalimat opini *K*. Sehingga algoritma ekstraksi fitur emosi dapat digambarkan dengan alur sebagai berikut [9].

- a. Mulai
- b. Jenis Emosi $E = [anger, anticipation, fear, surprise, sadness, joy, trust, and disgust] = [E_1, E_2, \dots, E_x]$. Jumlah emosi adalah *x*, dengan $x=7$
- c. Membaca *dataset Do*, dengan $Do = [K_1, K_2, K_3, \dots, K_n]$ dan *n*= banyaknya opini

- d. Inisialisasi `nextbaris = 1`
- e. Lakukan sebanyak `n` :
 1. membaca teks opini `K[nextbaris]`
 2. Konversikan teks opini `K` menjadi himpunan kata yang mandiri `W`, sehingga `K=[W1,W2,..Wm]` dengan `m` adalah banyaknya kata pada setiap opini `K[nextbaris]`
 3. Periksa setiap kata mandiri `W` tersebut yang berasosiasi dengan tipe emosi yang terdapat pada pustaka NRC (EmoLex)
 4. Setiap tipe emosi yang terkandung pada opini tersebut dicatat akumulasinya menjadi skor untuk setiap tipe emosi.
 5. Seleksi emosi yang tertinggi skor-nya.
 6. Dalam membedakan emosi yang tinggi dan rendah, skor setiap emosi dibagi dengan skor emosi tertinggi.
 - a. Apabila hasil pembagian > 0 , emosi tersebut diberi label bobot 1.
 - b. Apabila hasil pembagian $= 0$, emosi tersebut diberi label bobot 0.
 7. Tentukan `nextbaris++`
- f. Apabila `nextbaris ≤ jumlahrecord`, kembali ke langkah e.
- g. Selesai.

3. HASIL DAN PEMBAHASAN

Hasil dari proses data *crawling* diperoleh 2000 *tweets* yang terkini, berbahasa Inggris dan yang terpopuler yang berasosiasi dengan *keyword* "sinovac". Gambar 2 adalah cuplikan *source code* dari fungsi pada Rstudio untuk mengunduh (*scrapping*) data *tweet* secara online dari media Twitter.

```
#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 = ''))
```

Gambar 2. Kode Pengunduhan Data opini (*Tweets*)

Selanjutnya dalam setiap opini, ukuran setiap huruf diubah menjadi huruf kecil agar dapat mengantisipasi dualisme perhitungan jenis kata (*term*) jika kata yang sama namun ditulis dengan ukuran huruf yang berbeda. Ini dikarenakan analisis sentimen termasuk metode yang bersifat *case sensitive*. Kata 'Happy' dan 'happy' jika dibiarkan tertulis demikian, maka waktu untuk proses analisis sentimen akan menjadi lebih lambat karena adanya dualisme perhitungan pada kata yang bermakna sama (*happy*) [4]. Sebagaimana pada Gambar tersebut, dilakukan pemanggilan fungsi `x` untuk melakukan *case folding* atau konversi huruf-huruf dokumen kedalam huruf kecil, agar tidak ditemukan kata-kata yang sama namun diidentifikasi berbeda akibat kemunculan huruf capital pada kata-kata tersebut.

Hasil pengunduhan data melalui proses *crawling* akan terlihat sebagai kumpulan data (*dataset*) opini dimana tiap opini memiliki nomor atau *id record* 1 hingga 2000 opini

seperti pada Gambar 3. Hasil kumpulan opini ini masih mengandung beberapa komponen, tanda baca, dan karakter khas dari *tweet* yang perlu untuk dihapuskan agar kumpulan opini tersebut bersih dan setiap kata dalam setiap *record* opini menjadi token atau kata dasar. Pengubahan menjadi token ini akan lebih memudahkan untuk pencocokan kata dengan kumpulan *term* pada kamus leksikon ketika Naïve Bayes menghitung probabilitas kemunculan kata (*term*) yang sesuai dengan *term* pada kamus Lexicon [3].

```
"", "x"
"1", "CP Group has played down its link to Sinovac: It does hold a minority stake in a Hong
"2", "2 million doses of the Covid-19 vaccine Sinovac reach Pakistan https://t.co/m3LAVd9pL
"3", "Sinovac was the available vaccine! Remember that if your doctor says you have no spec
"4", "Thailand sticks with Sinovac vaccine after cases of 'stroke-like'...
https://t.co/BNF4tL5tOL via @MailOnline"
"5", "@gerry0514 @rapplerdotcom Actually, the survey has basis because Sinovac uses the old
"6", "RT @williedfw: You spent 5 years finding faults, calling the admin extortionist, fab
"7", "QUOTE UNQUOTE: "Another issue is people vaccinated with Sinovac will not be permitte
"8", "@erocq My Dad who is 69 yrs old got sinovac. Okay naman. Usual side effect na bumigat
"9", "<U+0001F9EA> Vaccines used globally
Sinovac:
<U+0001F1E6><U+0001F1F1> AL <U+0001F1E6><U+0001F1F2> AM <U+0001F1E6><U+0001F1FF> AZ <U+000
<U+0001F1E7><U+0001F1F7> BR <U+0001F1E8><U+0001F1F1> CL <U+0001F1E8><U+0001F1F3> CN <U+000
<U+0001F1E9><U+0001F1F4> DO <U+0001F1EA><U+0001F1E8> EC <U+0001F1EA><U+0001F1EC> EG <U+000
"10", "@zryanverse this is lazy reporting from Ars that you did not recognize. It doesn't s
"11", "@videobabe79 Rn for young peeps in my country it's just AZ.. VB my country gomen is
https://t.co/RvzFoh5tIS"
"12", "@RenHu2 I know a lot of people who prefer Sinovac over Pfizer (or Moderna) because t
"13", "RT @aidilarazak: Not sure why some say it's discrimination against single ppl.
```

Gambar 3. Data (*Tweets*) hasil *crawling* (pengunduhan opini)

Untuk menghapus tanda-tanda baca, dan URL pada setiap opini dilakukan dengan memanggil sejumlah baris fungsi pada *library* Twitter sebagaimana dalam Gambar 4.

```
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+0001F1E6>|<U+0001F1E7>|<U+0001F1E8>|<U+0001F1E9>|<U+0001F1EA>|<U+0001F1EB>|<U+0001F1EC>|<U+0001F1ED>|<U+0001F1EE>|<U+0001F1EF>|<U+0001F1F0>|<U+0001F1F1>|<U+0001F1F2>|<U+0001F1F3>|<U+0001F1F4>|<U+0001F1F5>|<U+0001F1F6>|<U+0001F1F7>|<U+0001F1F8>|<U+0001F1F9>|<U+0001F1FA>|<U+0001F1FB>|<U+0001F1FC>|<U+0001F1FD>|<U+0001F1FE>|<U+0001F1FF>|<U+0001F100>|<U+0001F101>|<U+0001F102>|<U+0001F103>|<U+0001F104>|<U+0001F105>|<U+0001F106>|<U+0001F107>|<U+0001F108>|<U+0001F109>|<U+0001F10A>|<U+0001F10B>|<U+0001F10C>|<U+0001F10D>|<U+0001F10E>|<U+0001F10F>|<U+0001F110>|<U+0001F111>|<U+0001F112>|<U+0001F113>|<U+0001F114>|<U+0001F115>|<U+0001F116>|<U+0001F117>|<U+0001F118>|<U+0001F119>|<U+0001F11A>|<U+0001F11B>|<U+0001F11C>|<U+0001F11D>|<U+0001F11E>|<U+0001F11F>|<U+0001F120>|<U+0001F121>|<U+0001F122>|<U+0001F123>|<U+0001F124>|<U+0001F125>|<U+0001F126>|<U+0001F127>|<U+0001F128>|<U+0001F129>|<U+0001F12A>|<U+0001F12B>|<U+0001F12C>|<U+0001F12D>|<U+0001F12E>|<U+0001F12F>|<U+0001F130>|<U+0001F131>|<U+0001F132>|<U+0001F133>|<U+0001F134>|<U+0001F135>|<U+0001F136>|<U+0001F137>|<U+0001F138>|<U+0001F139>|<U+0001F13A>|<U+0001F13B>|<U+0001F13C>|<U+0001F13D>|<U+0001F13E>|<U+0001F13F>|<U+0001F140>|<U+0001F141>|<U+0001F142>|<U+0001F143>|<U+0001F144>|<U+0001F145>|<U+0001F146>|<U+0001F147>|<U+0001F148>|<U+0001F149>|<U+0001F14A>|<U+0001F14B>|<U+0001F14C>|<U+0001F14D>|<U+0001F14E>|<U+0001F14F>|<U+0001F150>|<U+0001F151>|<U+0001F152>|<U+0001F153>|<U+0001F154>|<U+0001F155>|<U+0001F156>|<U+0001F157>|<U+0001F158>|<U+0001F159>|<U+0001F15A>|<U+0001F15B>|<U+0001F15C>|<U+0001F15D>|<U+0001F15E>|<U+0001F15F>|<U+0001F160>|<U+0001F161>|<U+0001F162>|<U+0001F163>|<U+0001F164>|<U+0001F165>|<U+0001F166>|<U+0001F167>|<U+0001F168>|<U+0001F169>|<U+0001F16A>|<U+0001F16B>|<U+0001F16C>|<U+0001F16D>|<U+0001F16E>|<U+0001F16F>|<U+0001F170>|<U+0001F171>|<U+0001F172>|<U+0001F173>|<U+0001F174>|<U+0001F175>|<U+0001F176>|<U+0001F177>|<U+0001F178>|<U+0001F179>|<U+0001F17A>|<U+0001F17B>|<U+0001F17C>|<U+0001F17D>|<U+0001F17E>|<U+0001F17F>|<U+0001F180>|<U+0001F181>|<U+0001F182>|<U+0001F183>|<U+0001F184>|<U+0001F185>|<U+0001F186>|<U+0001F187>|<U+0001F188>|<U+0001F189>|<U+0001F18A>|<U+0001F18B>|<U+0001F18C>|<U+0001F18D>|<U+0001F18E>|<U+0001F18F>|<U+0001F190>|<U+0001F191>|<U+0001F192>|<U+0001F193>|<U+0001F194>|<U+0001F195>|<U+0001F196>|<U+0001F197>|<U+0001F198>|<U+0001F199>|<U+0001F19A>|<U+0001F19B>|<U+0001F19C>|<U+0001F19D>|<U+0001F19E>|<U+0001F19F>|<U+0001F1A0>|<U+0001F1A1>|<U+0001F1A2>|<U+0001F1A3>|<U+0001F1A4>|<U+0001F1A5>|<U+0001F1A6>|<U+0001F1A7>|<U+0001F1A8>|<U+0001F1A9>|<U+0001F1AA>|<U+0001F1AB>|<U+0001F1AC>|<U+0001F1AD>|<U+0001F1AE>|<U+0001F1AF>|<U+0001F1B0>|<U+0001F1B1>|<U+0001F1B2>|<U+0001F1B3>|<U+0001F1B4>|<U+0001F1B5>|<U+0001F1B6>|<U+0001F1B7>|<U+0001F1B8>|<U+0001F1B9>|<U+0001F1BA>|<U+0001F1BB>|<U+0001F1BC>|<U+0001F1BD>|<U+0001F1BE>|<U+0001F1BF>|<U+0001F1C0>|<U+0001F1C1>|<U+0001F1C2>|<U+0001F1C3>|<U+0001F1C4>|<U+0001F1C5>|<U+0001F1C6>|<U+0001F1C7>|<U+0001F1C8>|<U+0001F1C9>|<U+0001F1CA>|<U+0001F1CB>|<U+0001F1CC>|<U+0001F1CD>|<U+0001F1CE>|<U+0001F1CF>|<U+0001F1D0>|<U+0001F1D1>|<U+0001F1D2>|<U+0001F1D3>|<U+0001F1D4>|<U+0001F1D5>|<U+0001F1D6>|<U+0001F1D7>|<U+0001F1D8>|<U+0001F1D9>|<U+0001F1DA>|<U+0001F1DB>|<U+0001F1DC>|<U+0001F1DD>|<U+0001F1DE>|<U+0001F1DF>|<U+0001F1E0>|<U+0001F1E1>|<U+0001F1E2>|<U+0001F1E3>|<U+0001F1E4>|<U+0001F1E5>|<U+0001F1E6>|<U+0001F1E7>|<U+0001F1E8>|<U+0001F1E9>|<U+0001F1EA>|<U+0001F1EB>|<U+0001F1EC>|<U+0001F1ED>|<U+0001F1EE>|<U+0001F1EF>|<U+0001F1F0>|<U+0001F1F1>|<U+0001F1F2>|<U+0001F1F3>|<U+0001F1F4>|<U+0001F1F5>|<U+0001F1F6>|<U+0001F1F7>|<U+0001F1F8>|<U+0001F1F9>|<U+0001F1FA>|<U+0001F1FB>|<U+0001F1FC>|<U+0001F1FD>|<U+0001F1FE>|<U+0001F1FF>|<U+0001F100>|<U+0001F101>|<U+0001F102>|<U+0001F103>|<U+0001F104>|<U+0001F105>|<U+0001F106>|<U+0001F107>|<U+0001F108>|<U+0001F109>|<U+0001F10A>|<U+0001F10B>|<U+0001F10C>|<U+0001F10D>|<U+0001F10E>|<U+0001F10F>|<U+0001F110>|<U+0001F111>|<U+0001F112>|<U+0001F113>|<U+0001F114>|<U+0001F115>|<U+0001F116>|<U+0001F117>|<U+0001F118>|<U+0001F119>|<U+0001F11A>|<U+0001F11B>|<U+0001F11C>|<U+0001F11D>|<U+0001F11E>|<U+0001F11F>|<U+0001F120>|<U+0001F121>|<U+0001F122>|<U+0001F123>|<U+0001F124>|<U+0001F125>|<U+0001F126>|<U+0001F127>|<U+0001F128>|<U+0001F129>|<U+0001F12A>|<U+0001F12B>|<U+0001F12C>|<U+0001F12D>|<U+0001F12E>|<U+0001F12F>|<U+0001F130>|<U+0001F131>|<U+0001F132>|<U+0001F133>|<U+0001F134>|<U+0001F135>|<U+0001F136>|<U+0001F137>|<U+0001F138>|<U+0001F139>|<U+0001F13A>|<U+0001F13B>|<U+0001F13C>|<U+0001F13D>|<U+0001F13E>|<U+0001F13F>|<U+0001F140>|<U+0001F141>|<U+0001F142>|<U+0001F143>|<U+0001F144>|<U+0001F145>|<U+0001F146>|<U+0001F147>|<U+0001F148>|<U+0001F149>|<U+0001F14A>|<U+0001F14B>|<U+0001F14C>|<U+0001F14D>|<U+0001F14E>|<U+0001F14F>|<U+0001F150>|<U+0001F151>|<U+0001F152>|<U+0001F153>|<U+0001F154>|<U+0001F155>|<U+0001F156>|<U+0001F157>|<U+0001F158>|<U+0001F159>|<U+0001F15A>|<U+0001F15B>|<U+0001F15C>|<U+0001F15D>|<U+0001F15E>|<U+0001F15F>|<U+0001F160>|<U+0001F161>|<U+0001F162>|<U+0001F163>|<U+0001F164>|<U+0001F165>|<U+0001F166>|<U+0001F167>|<U+0001F168>|<U+0001F169>|<U+0001F16A>|<U+0001F16B>|<U+0001F16C>|<U+0001F16D>|<U+0001F16E>|<U+0001F16F>|<U+0001F170>|<U+0001F171>|<U+0001F172>|<U+0001F173>|<U+0001F174>|<U+0001F175>|<U+0001F176>|<U+0001F177>|<U+0001F178>|<U+0001F179>|<U+0001F17A>|<U+0001F17B>|<U+0001F17C>|<U+0001F17D>|<U+0001F17E>|<U+0001F17F>|<U+0001F180>|<U+0001F181>|<U+0001F182>|<U+0001F183>|<U+0001F184>|<U+0001F185>|<U+0001F186>|<U+0001F187>|<U+0001F188>|<U+0001F189>|<U+0001F18A>|<U+0001F18B>|<U+0001F18C>|<U+0001F18D>|<U+0001F18E>|<U+0001F18F>|<U+0001F190>|<U+0001F191>|<U+0001F192>|<U+0001F193>|<U+0001F194>|<U+0001F195>|<U+0001F196>|<U+0001F197>|<U+0001F198>|<U+0001F199>|<U+0001F19A>|<U+0001F19B>|<U+0001F19C>|<U+0001F19D>|<U+0001F19E>|<U+0001F19F>|<U+0001F1A0>|<U+0001F1A1>|<U+0001F1A2>|<U+0001F1A3>|<U+0001F1A4>|<U+0001F1A5>|<U+0001F1A6>|<U+0001F1A7>|<U+0001F1A8>|<U+0001F1A9>|<U+0001F1AA>|<U+0001F1AB>|<U+0001F1AC>|<U+0001F1AD>|<U+0001F1AE>|<U+0001F1AF>|<U+0001F1B0>|<U+0001F1B1>|<U+0001F1B2>|<U+0001F1B3>|<U+0001F1B4>|<U+0001F1B5>|<U+0001F1B6>|<U+0001F1B7>|<U+0001F1B8>|<U+0001F1B9>|<U+0001F1BA>|<U+0001F1BB>|<U+0001F1BC>|<U+0001F1BD>|<U+0001F1BE>|<U+0001F1BF>|<U+0001F1C0>|<U+0001F1C1>|<U+0001F1C2>|<U+0001F1C3>|<U+0001F1C4>|<U+0001F1C5>|<U+0001F1C6>|<U+0001F1C7>|<U+0001F1C8>|<U+0001F1C9>|<U+0001F1CA>|<U+0001F1CB>|<U+0001F1CC>|<U+0001F1CD>|<U+0001F1CE>|<U+0001F1CF>|<U+0001F1D0>|<U+0001F1D1>|<U+0001F1D2>|<U+0001F1D3>|<U+0001F1D4>|<U+0001F1D5>|<U+0001F1D6>|<U+0001F1D7>|<U+0001F1D8>|<U+0001F1D9>|<U+0001F1DA>|<U+0001F1DB>|<U+0001F1DC>|<U+0001F1DD>|<U+0001F1DE>|<U+0001F1DF>|<U+0001F1E0>|<U+0001F1E1>|<U+0001F1E2>|<U+0001F1E3>|<U+0001F1E4>|<U+0001F1E5>|<U+0001F1E6>|<U+0001F1E7>|<U+0001F1E8>|<U+0001F1E9>|<U+0001F1EA>|<U+0001F1EB>|<U+0001F1EC>|<U+0001F1ED>|<U+0001F1EE>|<U+0001F1EF>|<U+0001F1F0>|<U+0001F1F1>|<U+0001F1F2>|<U+0001F1F3>|<U+0001F1F4>|<U+0001F1F5>|<U+0001F1F6>|<U+0001F1F7>|<U+0001F1F8>|<U+0001F1F9>|<U+0001F1FA>|<U+0001F1FB>|<U+0001F1FC>|<U+0001F1FD>|<U+0001F1FE>|<U+0001F1FF>|<U+0001F100>|<U+0001F101>|<U+0001F102>|<U+0001F103>|<U+0001F104>|<U+0001F105>|<U+0001F106>|<U+0001F107>|<U+0001F108>|<U+0001F109>|<U+0001F10A>|<U+0001F10B>|<U+0001F10C>|<U+0001F10D>|<U+0001F10E>|<U+0001F10F>|<U+0001F110>|<U+0001F111>|<U+0001F112>|<U+0001F113>|<U+0001F114>|<U+0001F115>|<U+0001F116>|<U+0001F117>|<U+0001F118>|<U+0001F119>|<U+0001F11A>|<U+0001F11B>|<U+0001F11C>|<U+0001F11D>|<U+0001F11E>|<U+0001F11F>|<U+0001F120>|<U+0001F121>|<U+0001F122>|<U+0001F123>|<U+0001F124>|<U+0001F125>|<U+0001F126>|<U+0001F127>|<U+0001F128>|<U+0001F129>|<U+0001F12A>|<U+0001F12B>|<U+0001F12C>|<U+0001F12D>|<U+0001F12E>|<U+0001F12F>|<U+0001F130>|<U+0001F131>|<U+0001F132>|<U+0001F133>|<U+0001F134>|<U+0001F135>|<U+0001F136>|<U+0001F137>|<U+0001F138>|<U+0001F139>|<U+0001F13A>|<U+0001F13B>|<U+0001F13C>|<U+0001F13D>|<U+0001F13E>|<U+0001F13F>|<U+0001F140>|<U+0001F141>|<U+0001F142>|<U+0001F143>|<U+0001F144>|<U+0001F145>|<U+0001F146>|<U+0001F147>|<U+0001F148>|<U+0001F149>|<U+0001F14A>|<U+0001F14B>|<U+0001F14C>|<U+0001F14D>|<U+0001F14E>|<U+0001F14F>|<U+0001F150>|<U+0001F151>|<U+0001F152>|<U+0001F153>|<U+0001F154>|<U+0001F155>|<U+0001F156>|<U+0001F157>|<U+0001F158>|<U+0001F159>|<U+0001F15A>|<U+0001F15B>|<U+0001F15C>|<U+0001F15D>|<U+0001F15E>|<U+0001F15F>
```

```
library(tm)
all_tweet<-dataengkap
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)
```

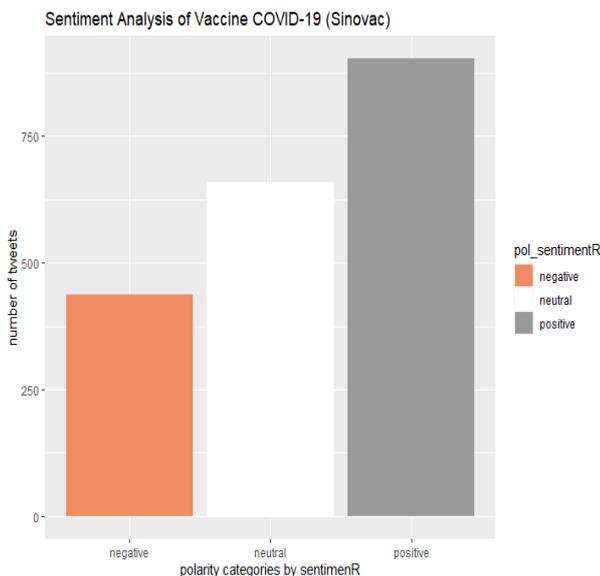
Gambar 5. Fungsi Penghapusan Stopword

Proses pembersihan data tersebut akan menjadikan setiap kata (*term*) sebagai *token* (kata yang bersih dari notasi atau imbuhan) seperti yang nampak pada Gambar 6.

```
"10", "lazi report ar recogn doesn t say explicit"
"11", "rn young peep countri s just az vb countri gomen selfi
"12", "know lot peopl prefer sinovac pfizer moderna scare nev
"13", "sure say s discrimin singl ppl marri ppl pregnant re d
"14", "komplek komun muhibbah kl may parent mysejahtera apmnt
"15", "hate bad news bear much trust despis hong kong govern
"16", "indonesia thus far receiv mil dose sinovac yesterday
"17", "spent year find fault call admin extortionist fabric
"18", "chines sinovac just packag differ name pr stunt"
"19", "differ sweden vaccin china sinovac sinovac base kill
"20", "anoth brand sinovac benefit gone away like will necess
"21", "begin hkgov cover risk vaccin never open tell yo sta
"22", "indonesia thus far receiv mil dose sinovac yesterday
"23", "woke peopl realli work china d pimp sinovac just say
"24", "indonesia thus far receiv mil dose sinovac yesterday
"25", "moment push sinovac authoris use think will allow ever
"26", "sinovac fact indonesia already vaccin health worker s
"27", "indonesia thus far receiv mil dose sinovac yesterday
"28", "liter say sinopharm articl tweet s sinovac"
"29", "doubt slot az sarawak will full later az go sinovac p
"30", "ll fulfil role ve bargain sinovac good vaccin ll low"
"31", "indonesia thus far receiv mil dose sinovac yesterday"
```

Gambar 6. Tweet Bersih hasil Cleaning dan Normalisasi

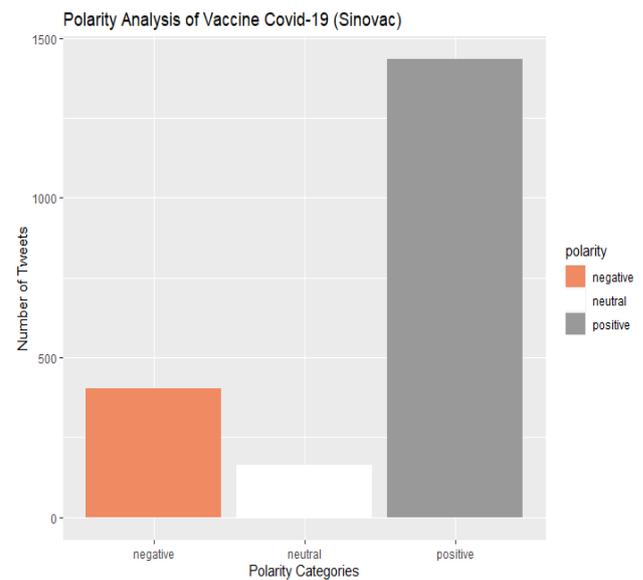
Dataset opini yang bersih selanjutnya diklasifikasikan berdasarkan kategori sentimennya menggunakan fungsi `sentiment_by()` dari pustaka `sentimentR` kemudian dihasilkan kelas sentimen (polaritas) berupa negatif, netral, dan positif pada Gambar 7.



Gambar 7. Hasil Klasifikasi Tweets dengan Fungsi `sentiment_by()`

Fungsi `sentiment_by()` dari `Library sentimentR` menjalankan klasifikasi opini dengan metode Valence Shifter Lexicon menghitung nilai polaritas dari setiap kata-kata dalam opini yang sesuai dengan kata-kata yang berasosiasi dengan kategori sentimen pada kamus `Lexicon`,

disamping itu juga memperhitungkan kata-kata yang merupakan Valence Shifter dengan nilai valensi menurut kamus `Lexicon` [3][19]. Klasifikasi dengan fungsi `sentiment_by()` memberikan hasil yang ditunjukkan pada Gambar 7 yakni sentimen masyarakat terhadap vaksin “sinovac” sebanyak 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebanyak 660 (33%). Selanjutnya, fungsi `classify_polarity()` dari `library sentiment` yang menggunakan metode Naïve Bayes juga diimplementasikan dalam penelitian ini. Hasil klasifikasi sentimen dari `library sentiment` (metode Naïve Bayes) seperti pada Gambar 8, yakni sebanyak 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) sentimen netral.

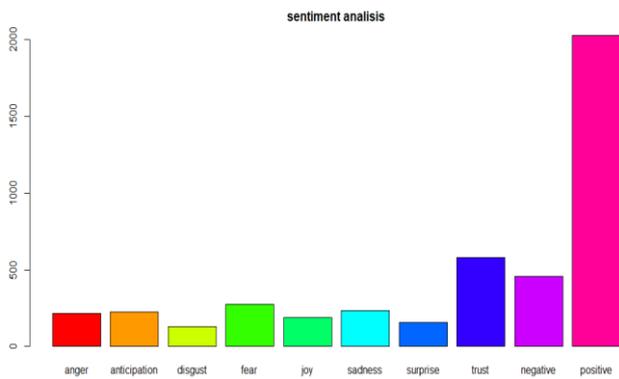


Gambar 8. Hasil Klasifikasi Menggunakan Fungsi `Classify_polarity`

Berdasarkan penelitian sebelumnya, `Library sentimentR` merupakan yang terbaik dalam mengklasifikasikan opini positif dan negatif [19]. Akan tetapi, jika dibandingkan dengan hasil yang nampak pada Gambar 7, ternyata hasil `library sentiment` (Naïve Bayes) memiliki jumlah polaritas (nilai sentimen) netral yang lebih sedikit dan polaritas positif yang lebih besar. Sehingga dapat dimungkinkan bahwa sebagian besar opini yang memiliki sentimen netral dari hasil klasifikasi oleh `library sentimentR` adalah sentimen positif. Terlihat dalam Gambar 7 maupun Gambar 8, hasil dari `library sentiment` (metode Naïve Bayes) lebih signifikan dalam membedakan sentimen positif dan negatif.

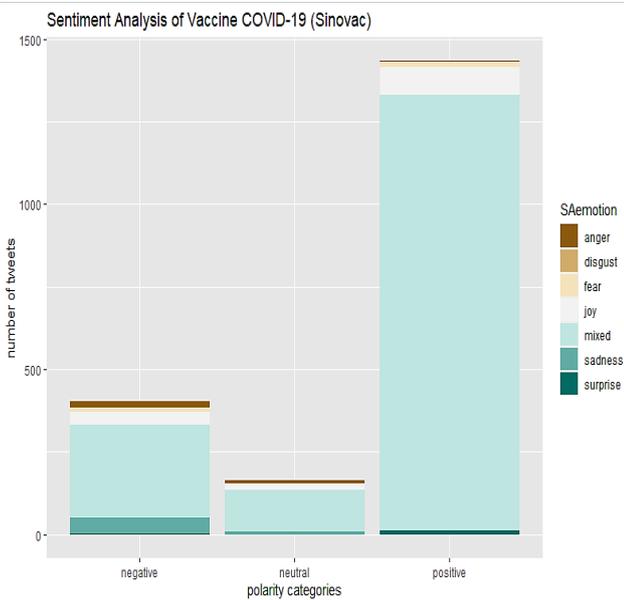
Sebelum mengetahui keterkaitan antara emosi dengan polaritas sentimen, maka dilakukannya klasifikasi emosi keseluruhan data opini menggunakan `library syuzhet` pada fungsi `get_nrc_sentiment` yang memberikan hasil pada Gambar 9. Fungsi `get_nrc_sentiment` juga memberikan hasil polaritas sentimen positif yang

secara signifikan lebih tinggi daripada sentimen negatif dan 8 kategori emosi lainnya.



Gambar 9. Hasil Klasifikasi Emosi Menggunakan Library Syuzhet

Library syuzhet dan library sentiment lebih cenderung sedikit menghasilkan polaritas netral daripada library sentimentR, dan hasil ini tentu berlawanan dengan penelitian sebelumnya [19]. Namun penelitian ini tidak berfokus pada keterkaitan emosi dan polaritas sentiment pada library syuzhet, karena akan lebih membahas perbandingan hasil klasifikasi sentimen yang dihasilkan library sentimentR dan library sentiment serta hubungan antara emosi dan sentimen berdasarkan hasil dari metode Naïve Bayes (library sentiment). Untuk melihat hubungan distribusi emosi dengan polaritas sentimen maka dilakukan visualisasi data dari library sentiment (metode Naïve Bayes) pada klasifikasi sentimen dan emosi sebagaimana pada Gambar 10.

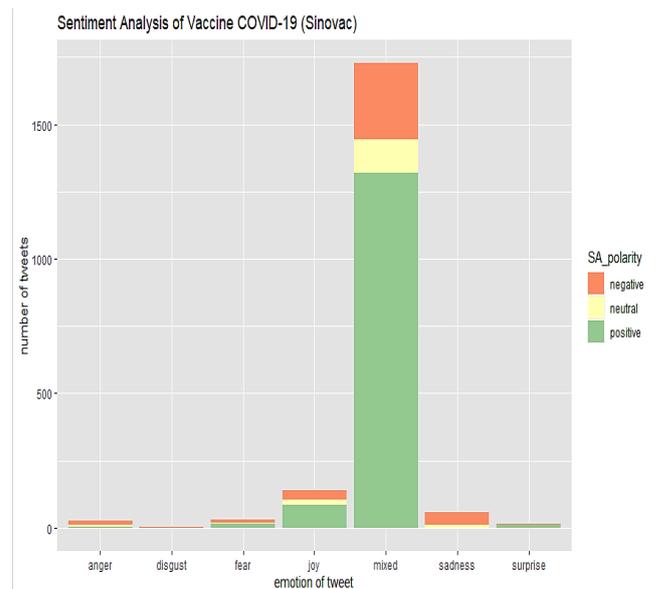


Gambar 10. Diagram Intensitas Emosi pada Setiap Kelas Sentimen

Gambar 10 menunjukkan bahwa setiap jenis polaritas sentimen mengandung jenis emosi mixed. Kategori emosi dianggap mixed apabila probabilitas term yang muncul di suatu opini memiliki skor tertinggi pada lebih dari satu jenis emosi. Pada Gambar 10 juga terlihat polaritas negatif

paling dipengaruhi oleh emosi anger, disgust, joy, dan sadness. Pada polaritas sentimen negatif, proporsi joy dan sadness hampir berimbang, sehingga dapat diketahui bahwa proporsi anger dan disgust sangat mempengaruhi polaritas negatif. Pada polaritas positif, proporsi joy adalah yang terbesar daripada proporsi joy pada jenis polaritas negatif dan netral. Disamping itu, emosi mixed lebih banyak pada polaritas positif. Untuk memperjelas keterkaitan emosi dan polaritas (nilai sentimen), maka dilakukan visualisasi grafik yang menunjukkan intensitas sentimen berdasarkan jenis emosi seperti pada Gambar 11. Pada Gambar 11, dapat terlihat bahwa jenis emosi:

- anger berasosiasi dengan sentimen negatif,
- disgust berasosiasi dengan sentimen negatif,
- fear berasosiasi dengan sentimen negatif dan positif
- joy cenderung berasosiasi dengan positif, sedikit berasosiasi dengan negatif dan netral
- mixed cenderung berasosiasi dengan positif, sedikit dengan negatif dan netral
- sadness cenderung berasosiasi dengan negatif, dan sedikit berasosiasi dengan netral
- surprise sedikit berasosiasi dengan negatif dan sedikit berasosiasi dengan positif



Gambar 11. Diagram Intensitas Sentimen pada Setiap Emosi

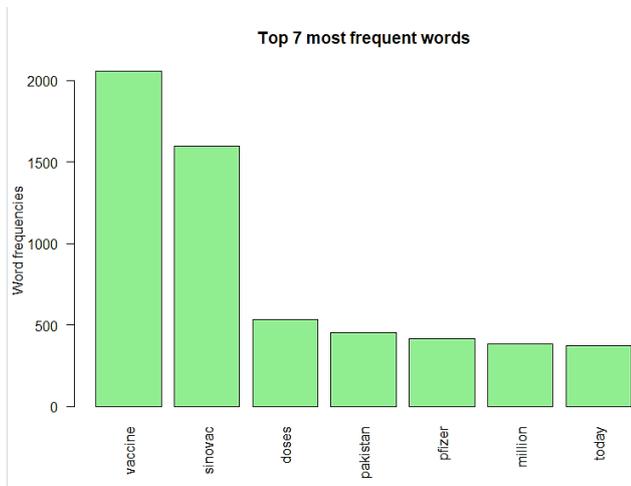
Gambar 12 menunjukkan hasil klasifikasi dengan fungsi sentiment_by() dari Library sentimentR terlihat bahwa beberapa cuplikan opini dengan nilai polaritas tidak sama dengan nol (tidak netral), lebih kecil dari nol (negatif), dan positif (lebih besar dari nol). Kategori polaritas tersebut dipengaruhi dengan proporsi kemunculan kata kunci yang mengarah pada kategori sentimen tertentu. Sebagai contoh pada sampel opini yang berpolaritas negatif terdapat kata-kata seperti : minority, fault, lazy, corrupted, hate, dan sebagainya. Sedangkan pada opini berpolaritas positif terdapat kata-kata kunci yang menjadi

ciri dari opini positif seperti *good*, *effective*, dan sebagainya. Pada opini dengan polaritas sentimen netral sangat jarang terdapat penggunaan kata-kata kunci opini negatif maupun positif. Semakin tingginya kemunculan kata-kata ekspresi ungkapan emosi dari opini yang mengarah ke sedih, marah, ataupun bahagia maka akan mempengaruhi jenis sentimennya yang cenderung mengarah pada polaritas positif atau ke polaritas negatif [5]. Pada penelitian ini, berdasarkan *library sentiment*, *tweet* dengan polaritas sentimen positif memiliki persentase paling tinggi dibandingkan polaritas negatif maupun netral untuk kata kunci “sinovac”.

3. X Numeric	4. text Nominal	5. ave_sentiment Numeric
1.0	cp group has played down its link to sinovac it does hold a minority stake in a hong kong company that in turn h	-0.951031036...
2.0	million doses of the covid vaccine sinovac reach pakistan	0.0
3.0	sinovac was the available vaccine remember that if your doctor says you have no special requirements the best...	-0.114707866...
4.0	thailand sticks with sinovac vaccine after cases of stroke like	0.079056941...
5.0	actually the survey has basis because sinovac uses the old vaccine technology the other	0.0267261241...
6.0	you spent years finding faults calling the admin extortionist fabricating palit ulo reporting to us officials bias ag	-0.353553390...
7.0	quote unquote another issue is people vaccinated with sinovac will not be permitted to enter many western cou...	-0.235702260...
8.0	my dad who is yrs old got sinovac okay naman usual side effect na bumigat braso parang lalagnatin which	0.0917662935...
9.0	vaccines used globally sinovac u f e u f f a l u f e u f f a m u f e u f f a z u f e u f e u f e b a u f e u f f o r u f e u f...	0.0
10.0	this is lazy reporting from ars that you did not recognize it doesn't say both of them explicitly the	-0.167705098...
11.0	in for young peeps in my country it's just az vjo my country gomen is so selfish corrupted u f u f they	-0.326598632...
12.0	i know a lot of people who prefer sinovac over pfizer or moderna because they are scared of the newer tec	-0.087287155...
13.0	not sure why some say it's discrimination against single ppl not all married ppl are pregnant if you're an adult	-0.106600358...
14.0	was at kompleks komunitas muhibbah kl may for parent mseyehahtera apmnt how come many young ppl getting u...	0.0942809041...
15.0	i hate to be a bad news bear but as much as i myself do not trust and despise the hong kong government my f...	-0.472455591...
16.0	indonesia have thus far received mil doses of sinovac yesterday they received another mil doses why they can...	0.2752988806...
17.0	you spent years finding faults calling the admin extortionist fabricating palit ulo reporting to us officials	-0.1875
18.0	this is chinese sinovac just packaged here under a different name pr stunt only it is	-0.05
19.0	the difference between sweden vaccine and china sinovac sinovac based on killed sarscov viruses	-0.133630620...
20.0	another brand after sinovac if its benefits have gone away most likely will that be necessarily on	0.2425356250...
21.0	from the very beginning hgov covered the risk of vaccination they never openly tell those over yo on these sta	-0.436033255...
22.0	indonesia have thus far received mil doses of sinovac yesterday they received another mil doses why they can...	0.2752988806...
23.0	if the woke people really would be working for china they'd be pimping sinovac just saying u f u f	-0.307009031...
24.0	indonesia have thus far received mil doses of sinovac yesterday they received another mil doses why they can...	0.2752988806...
25.0	at the moment they are pushing for sinovac to be authorised for use i think it will be allowed eventually	0.0
26.0	sinovac facts indonesia already vaccinated their health workers since jan the study conducted on more than peo	0.1576481562...
27.0	indonesia have thus far received mil doses of sinovac yesterday they received another mil doses why they can...	0.2752988806...

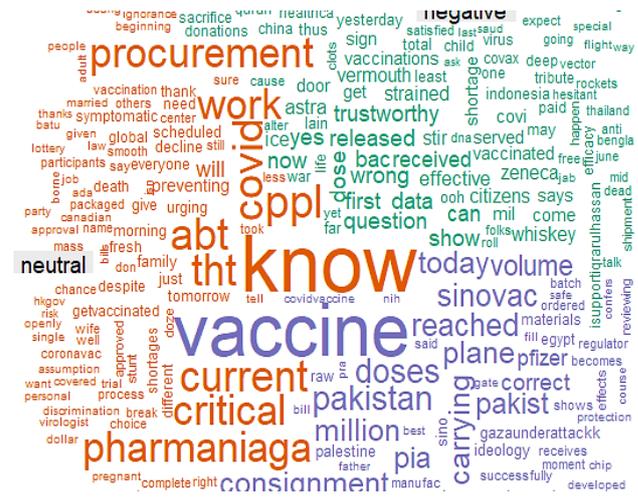
Gambar 12. Data Opini beserta Nilai Polaritas Sentimen dari Fungsi sentiment_by pada *Library sentiment*

Polaritas sentimen positif tertinggi dipengaruhi tingginya frekuensi kata-kata (*term*) yang muncul sebagaimana visualisasi *word frequency* pada Gambar 13. Kemunculan kata *Vaccine*, *sinovac*, *doses*, dan *Pakistan* mendominasi pada data *tweet* yang terkumpul. Fenomena tersebut memberikan sentimen positif karena munculnya vaksin Sinovac memberikan harapan baru untuk melawan dampak pandemi covid-19 walaupun seiring perkembangan waktu dan munculnya mutasi virus menyebabkan efisiensi vaksin menjadi isu utama yang perlu dikaji lebih lanjut.



Gambar 13. Most of Frequent Words (Diagram intensitas Kemunculan Kata terbanyak)

Berdasarkan perolehan kata terbanyak pada seluruh data *tweet*, dapat direpresentasikan dalam suatu *worldcloud* yang menunjukkan kata-kata terbanyak yang muncul berdasarkan polaritas sentimen menurut *library sentiment* (metode Naïve Bayes) pada fungsi *classify_polarity()* sebagaimana pada Gambar 14. Semakin besar suatu kata muncul menunjukkan semakin tingginya frekuensi kemunculan kata tersebut dalam kelas polaritas sentimen tertentu.



Gambar 14. *Wordcloud* dari kata-kata terbanyak berdasarkan kelas sentimen (polaritasnya)

Worldcloud pada Gambar 14 menunjukkan kata-kata yang berwarna hijau merupakan yang terbanyak muncul pada kelas sentimen negatif, seperti “Zeneca”, ”whiskey”, ”virus”, “war”, “yet fair”, “hesitant”, dan sebagainya. Kata-kata yang berwarna merah merupakan yang terbanyak muncul pada kelas sentimen netral, seperti “know”, ”procurement”, “getvaccinated” dan sebagainya. Sedangkan kata-kata yang berwarna ungu merupakan yang terbanyak muncul pada kelas sentimen positif, seperti “vaccine”, “sinovac”, “Pfizer”, “correct” dan sebagainya.

4. KESIMPULAN

Berdasarkan hasil analisis sentimen melalui klasifikasi dengan metode Naïve Bayes (*library sentiment*) telah mendeteksi sentimen positif dengan intensitas terbanyak, dan sentimen netral dengan intensitas paling sedikit dibandingkan dengan metode Valence Shifter Lexicon pada *library sentiment*. Penelitian ini menunjukkan hasil bahwa Naïve Bayes mengenali sebanyak 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) sentimen netral. Sedangkan Valence Shifter Lexicon mengenali sebanyak 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebesar 660 (33%). Metode Naïve Bayes juga mengenali emosi dengan jumlah terbanyak yakni emosi campuran (*mixed*) 1727(86,35%) dan emosi *joy* (gembira) sebanyak 141 (7,05%).

Baik metode Naïve Bayes maupun Valence Shifter Lexicon telah berhasil mengklasifikasikan opini masyarakat terkait vaksin Sinovac dengan hasil pada kedua metode yang menunjukkan sentimen publik masih cenderung positif. Hal ini mengindikasikan pada awal tahun 2021 informasi vaksin Sinovac menginspirasi harapan positif untuk menghadapi pandemi covid-19.

Klasifikasi emosi terhadap data *tweet* dalam penelitian ini menunjukkan kategori emosi *joy*, dan *mixed* memiliki persentasi terbanyak yang mengandung polaritas sentimen positif. Emosi sangat ditentukan oleh *term* atau kata yang muncul dan cocok dengan kumpulan *term* dalam lexicon. Sehingga satu kata dapat berasosiasi pada banyak kategori emosi. Untuk penelitian kedepan, akan lebih baik jika data *tweet* dengan emosi *mixed* diklasifikasikan dengan data-data *tweet* yang berkategori emosi spesifik dengan pengujian akurasi berdasarkan pelabelan manual, sehingga diharapkan dapat mengurangi persentasi kategori emosi *mixed*.

Ucapan Terima Kasih

Ucapan terima kasih Penulis sampaikan kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Pembangunan Nasional Veteran Yogyakarta, atas dukungan dan dana dalam menjalankan penelitian ini melalui program hibah internal LPPM UPN Veteran Yogyakarta tahun 2021.

DAFTAR PUSTAKA

- [1] 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.
- [2] 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.
- [3] B. M. Akbar, A. T. Akbar, and R. Husaini, "Classification of Sentiments on Twitter Opinions with The Keyword Sinovac Using Naïve Bayes," *Pros. Semin. Nas. Inform.*, vol. 1, no. 1, pp. 161–172, 2021, [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/6070/3929>.
- [4] 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>.
- [5] 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.
- [6] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [7] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*, 2010, pp. 627–666.
- [8] 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>.
- [9] A. S. Aribowo and S. Khomsah, "Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19)," *Telematika*, vol. 18, no. 1, p. 49, 2021, doi: 10.31315/telematika.v18i1.4341.
- [10] 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.
- [11] A. T. Akbar, R. Husaini, B. M. Akbar, and S. Saifullah, "A proposed method for handling an imbalance data in classification of blood type based on Myers-Briggs type indicator," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 4, pp. 276–283, 2020, doi: 10.14710/jtsiskom.2020.13625.
- [12] L. S.Katore and J. S. U. J.S.Umale, "Comparative Study of Recommendation Algorithms and Systems using WEKA," *Int. J. Comput. Appl.*, vol. 110, no. 3, pp. 14–17, 2015, doi: 10.5120/19295-0731.
- [13] 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>.
- [14] F. Ratnawati, "Implementasi Algoritma Naïve 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.
- [15] Andros, D. Prawita, J. Karsten, and M. Vinandar, "Perbandingan Algoritma Pendeteksian Spam," *J.*

- Teknologi Terpadu*, vol. 1, no. 1, pp. 1–5, 2015.
- [16] N. Saputra, “(Sentiment Analisis With Lexicon Preprocessing),” *Din. Inform.*, vol. 7, no. 1, pp. 45–57, 2019.
- [17] I. Rish, “An Empirical Study of the Naïve Bayes Classifier An empirical study of the Naïve Bayes classifier,” no. January 2001, 2014, [Online]. Available: https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier/link/00b7d52dc3ccd8d692000000/download.
- [18] J. Jtik, J. Teknologi, R. T. Aldisa, and M. A. Abdullah, “Analisis Sentimen Mengenai Vaksin Sinovac di Media Sosial Twitter Menggunakan Metode Naïve bayes Classification,” vol. 6, no. 3, pp. 1–5, 2022.
- [19] 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>.
- [20] E. Riloff and J. Wiebe, “Learning extraction patterns for subjective expressions,” pp. 105–112, 2003, doi: 10.3115/1119355.1119369.
- [21] C. Strapparava and A. Valitutti, “WordNet-Affect: An affective extension of WordNet,” *Proc. 4th Int. Conf. Lang. Resour. Eval. Lr. 2004*, pp. 1083–1086, 2004.
- [22] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016.