

## ***Analysis of Sentiments and Emotions about Sinovac Vaccine Using Naive Bayes***

Analisis Emosi dan Sentimen tentang Vaksin Sinovac Menggunakan Naive Bayes

**Bagus Muhammad Akbar<sup>1</sup>, Ahmad Taufiq Akbar<sup>2</sup>, Rochmat Husaini<sup>3</sup>**

Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

<sup>1</sup>bagusmuhammadakbar@upnyk.ac.id, <sup>2</sup>ahmadtaufiq.akbar@upnyk.ac.id,

<sup>3</sup>husaini@upnyk.ac.id

### ***Article's Information / Informasi Artikel***

*Received: April 2022*

*Revised: May 2022*

*Accepted: June 2022*

*Published: June 2022*

### ***Abstract***

#### ***Purpose:***

*Many countries in the world have tried to control the impact of the COVID-19 pandemic through the use of vaccines. Sinovac vaccine is one of the popular vaccines that has been used in several countries including Indonesia. Since the presence of the Sinovac vaccine, public perceptions both in the field and on social media have increasingly emerged between agreeing and disagreeing with the vaccine. The perception of the world community on social media can be analyzed to determine the sentiment category and the emotional level of the community re-garding the acceptance of the Sinovac vaccine.*

#### ***Design/methodology/approach:***

*The analysis can be done through data mining that uses the Naive Bayes algorithm to calculate probabilities and statistics so that each opinion can be classified in the category of positive, negative, or neutral sentiment. In this study, the source of data analysis is public perception which contains the keyword "sinovac" from the twitter. Tests using the senti-ment, sentimentr, and syuzhet library show that positive sentiment is higher than negative and neutral. Negative sentiments are most influenced by the emotional level of sadness and anger. While positive sentiment is strongly influenced by the category of joy and mixed emotions. The mixed emotion category is more correspond to the positive sentiment.*

#### ***Findings/result:***

*Based on this research, the keyword sinovac tends to generate positive sentiment. Polarity affects Emotions, but not the other way around. Because it is seen that the value of polarity classification accuracy (with both libraries)*

---

*increases when the emotion feature is not included. While the value of accuracy in classifying emotions actually increases when the polarity feature is included.*

*Originality/value/state of the art:*

*The Naive Bayes method (library sentiment) and the valence shifter (library sentimentr) method used in sentiment analysis in this study indicate that positive sentiment is higher than neutral and negative. The result of the percentage of positive sentiment by the valence shifter method is lower than the Naive Bayes method. The Valence Shifter method tends to produce a smaller aggregate between the percentage results of positive sentiment compared to neutral and negative.*

### **Abstrak**

*Keywords: sentiment analysis, naïve bayes, vaccine*

Kata kunci: analisis sentimen, naïve bayes, vaksin

**Tujuan:**

Banyak negara di dunia telah berusaha mengendalikan dampak pandemi COVID-19 melalui penggunaan vaksin. vaksin sinovac merupakan salah satu vaksin populer yang telah digunakan di beberapa negara termasuk Indonesia. Sejak hadirnya vaksin sinovac, persepsi masyarakat baik di lapangan maupun di media sosial semakin muncul antara setuju dan tidak setuju dengan vaksin tersebut. Persepsi masyarakat dunia di media sosial dapat dianalisis untuk mengetahui kategori sentimen dan tingkat emosional masyarakat terhadap penerimaan vaksin Sinovac.

**Perancangan/metode/pendekatan:**

Analisis dapat dilakukan melalui data mining yang menggunakan algoritma Naive Bayes untuk menghitung probabilitas dan statistik sehingga setiap opini dapat diklasifikasikan dalam kategori sentimen positif, negatif, atau netral. Dalam penelitian ini, sumber analisis data adalah persepsi publik yang mengandung kata kunci “sinovac” dari twitter. Pengujian menggunakan sentimen, sentimen, dan *library* syuzhet menunjukkan bahwa sentimen positif lebih tinggi daripada negatif dan netral. Sentimen negatif paling dipengaruhi oleh tingkat emosional kesedihan dan kemarahan. Sedangkan sentimen positif sangat dipengaruhi oleh kategori senang dan emosi campur aduk. Kategori emosi campuran lebih sesuai dengan sentimen positif.

**Hasil:**

Klasifikasi emosi terhadap data *tweet* dalam penelitian ini menunjukkan bahwa kategori emosi kegembiraan, dan campuran memiliki persentase tertinggi yang mengandung polaritas sentimen positif. Berdasarkan penelitian ini, kata kunci sinovac cenderung memunculkan sentimen positif.

---

---

Polaritas mempengaruhi emosi, namun tidak sebaliknya. Karena terlihat bahwa nilai akurasi pada klasifikasi polaritas (dengan kedua *library*) telah meningkat ketika fitur emosi tidak diikuti. Sedangkan nilai akurasi pada klasifikasi emosi justru meningkat ketika fitur polaritas diikuti.

Keaslian/ *state of the art*:

Metode Naive Bayes (*library sentiment*) dan metode *Valence Shifter* (*library sentimentr*) yang digunakan dalam analisis sentimen pada penelitian ini menunjukkan bahwa sentimen positif lebih tinggi daripada netral dan negatif. Hasil persentase sentimen positif oleh metode *Valence Shifter* lebih rendah daripada metode Naive Bayes. Pada metode *Valence Shifter* cenderung menghasilkan agregat yang lebih kecil antara hasil persentase sentimen positif dibanding netral dan negatif.

---

## 1. Pendahuluan

Penanggulangan virus corona di awal tahun 2021 dengan menggunakan vaksin sudah mulai dilakukan oleh pemerintah di seluruh dunia [1]. Ini menimbulkan masalah baru tentang penerimaan vaksin ke publik [2]. Berbagai tanggapan di media sosial untuk menyampaikan aspirasi masyarakat tentang vaksin COVID-19 bermunculan, salah satunya melalui aplikasi twitter [3]. Jumlah pengguna *twitter* meningkat sangat pesat lebih dari 166 juta pengguna terdaftar mulai tahun 2020 [4]. Tingkah laku orang-orang yang terbiasa memposting kicauan tentang fenomena yang sedang hangat-hangatnya saat ini mulai muncul daya tariknya [1][5]. Oleh karena itu, untuk mengetahui emosi pengguna media sosial, dilakukan analisis sentimen untuk meninjau opini publik dengan melakukan data mining terhadap kumpulan kutipan pengguna di *twitter* tentang vaksin COVID-19 khususnya vaksin sinovac [2].

Analisis sentimen merupakan penggalian opini dari suatu teks menggunakan metode komputasi yang dapat berupa metode statistik dan *machine learning* untuk mengenali beberapa aspek yang terkandung dalam teks seperti subjektivitas, kategori sentimen, kategori emosi, pola sikap, karakter dan sebagainya [6]. Analisis sentimen juga dapat dikembangkan menjadi metode yang secara otomatis dapat mengekstraksi opini dari media sosial elektronik agar dapat dipahami apakah konteks dan sentimen dalam teks tersebut, sehingga akan bermanfaat untuk membendung konten yang tidak layak untuk disiarkan secara publik[7][8]. Emosi merupakan bagian dari aspek kecerdasan manusia dan memiliki peran penting dalam pengambilan keputusan yang dibuat oleh manusia. Saat manusia berinteraksi dengan komputer, banyak hubungan emosional yang terjadi seringkali diabaikan padahal itu sangat penting dalam komunikasi manusia dalam kehidupan kita sehari-hari [5][9]. Ada enam emosi dasar dalam diri manusia, yaitu senang, sedih, takut, marah, terkejut, dan jijik [10]. Tidak dapat dipungkiri bahwa emosi memiliki peran penting dalam komunikasi antar manusia dalam kehidupan sehari-hari [11].

Penelitian tentang analisis sentimen telah dilakukan oleh beberapa peneliti sebelumnya. Penelitian penerapan *normalized poly kernel* pada *machine learning SVM* (*Support Vector Machine*) banyak digunakan dalam beberapa riset analisis sentimen seperti analisis respon pemerintah Indonesia dalam penanganan COVID-19 [3]. Penggunaan *Latent Dirichlet Allocation* (LDA) dalam mengelompokkan tweet topik yang akan diklasifikasikan

menggunakan metode *Valence Shifter* juga telah digunakan dalam penelitian untuk analisis sentimen tweet migrain [12]. Perbandingan pembelajaran mesin dalam analisis sentimen tweet Malayalam menyimpulkan bahwa hutan acak memiliki akurasi yang lebih baik daripada Naive Bayes dan SVM [13]. Pemrosesan bahasa alami menggunakan Naive Bayes menunjukkan akurasi yang baik dalam penelitian analisis sentimen vaksin COVID-19 di Filipina [1]. Penelitian sebelumnya menggunakan Vader Lexicon pada analisis sentimen, klasifikasi emosi dengan RNN, dan pemodelan topik dengan LDA menunjukkan bahwa hasil tertinggi adalah sentimen positif yang didominasi oleh emosi kegembiraan [14]. Dalam penelitian lain telah diuji beberapa metode analisis sentimen pada platform R yakni *library Rsentiment* yang tertinggi akurasinya, diikuti *library syuzhet*, *library meanr* dan *library Sentiment Analysis* [15]. Beberapa penelitian sebelumnya yang menggunakan R *Package* masih mengandalkan *library sentimentr*, sehingga pada penelitian ini ingin menggunakan *library sentimentr* dan *sentiment* dari R *Package* untuk menganalisis sentimen tentang vaksin Sinovac dari opini publik di *twitter*. Kami berharap penelitian kami dapat memberikan beberapa yaitu dapat memberikan informasi mengenai sentimen dan emosi publik terhadap vaksin COVID-19 sehingga pemerintah maupun masyarakat dapat menggunakan studi ini untuk membuat edukatif gambaran umum dan keputusan bijak tentang program vaksinasi.

## 2. Metode Penelitian

Penelitian diawali dengan pengumpulan data *tweet* yang menggunakan API *twitter* yang terdiri 5 komponen autentikasi yakni *consumer key*, *consumer secret key*, *access token key* dan *access secret key* sehingga peneliti mendapat hak akses untuk mengunduh data *tweet* [5][11]. Setelah data terkumpul, selanjutnya dilakukan *preprocessing*. Tahap *preprocessing* ini dengan proses *case folding* atau mengubah semua huruf teks menjadi huruf kecil (tidak kapital). Karena sistem analisis sentimen umumnya *case sensitif* sehingga jangan sampai terdapat kata-kata yang sama namun besar kecil hurufnya berbeda [16]. Kemudian penghapusan emoji, tanda baca dan melakukan tokenisasi dengan *tokenizer* sehingga diperoleh data opini yang telah bersih dan siap untuk dilakukan *opinion mining*.

Setiap data opini yang telah bersih tersebut kemudian diklasifikasikan dengan metode *Valence Shifter (library sentimentr)* untuk mendapatkan label berupa kategori sentimen, serta diklasifikasikan juga dengan metode Naive Bayes (*library sentiment*) untuk mendapatkan label berupa kategori emosi dan kategori sentimen. Dengan demikian pelabelan dilakukan oleh kedua metode tersebut berdasarkan leksikon pada *library*. Hasil dari kedua *library* ini akan memunculkan persentase sentimen dari kedua metode, dan persentase kategori emosi dari metode Naive Bayes. Pada tahap akhir adalah pengujian dan evaluasi dengan melakukan klasifikasi teks opini untuk melihat lebih jelas hubungan antara kategori emosi dan sentimen [17].

### 2.1. Sumber Data

Pada penelitian ini, dataset *tweet* diambil dari *twitter* dengan kata kunci "sinovac" menggunakan *tools* dari Rstudio yakni *library TwitteR*. Data *tweet* yang dikumpulkan sebanyak 2000 data yang merupakan data *tweet* yang terbaru dan paling populer sejak 1 Februari 2021[18][19].

### 2.2. Preprocessing

*Preprocessing* merupakan salah satu tahapan menghilangkan masalah yang dapat mengganggu hasil pengolahan data. Pada kasus klasifikasi dokumen yang menggunakan data bertipe teks,

*preprocessing* umumnya dilakukan dengan cara *case folding*, *filtering* (menghilangkan tanda baca), *tokenization*, *stopword removal*, *stemming*, dan sebagainya [11][3].

Data *tweet* yang telah diunduh ditampung dalam *file* berformat *csv* selanjutnya dilakukan pembersihan data dari tanda baca, karakter-karakter seperti *hashtag*, URL dan karakter lainnya yang tidak berguna dalam analisis sentimen. Pada data teks tersebut juga dilakukan *case folding* atau perubahan ke huruf kecil untuk menghindari ambiguitas pada analisis sentimen jika terdapat kata yang sama namun ditulis dengan huruf besar dan kecil. Selanjutnya data dilakukan tokenisasi menggunakan *tokenizer* pada *library* untuk memisahkan setiap kata dalam kalimat data teks kemudian *stemming* untuk mengembalikan kata ke bentuk dasarnya. [20][13][14].

Penghapusan *stopwords* dalam bahasa inggris pada penelitian ini menggunakan fungsi dari *library* *tm*. *Stopwords* adalah daftar kata yang mempunyai fungsi tetapi tidak mempunyai arti, dan *stopwords* ini tidak diprioritaskan dalam analisis sentimen [13]. Penelitian sebelumnya [21] telah menunjukkan bahwa pengujian analisis sentimen dari kumpulan opini yang tidak mengandung *stopwords* memiliki akurasi yang lebih tinggi daripada analisis sentimen pada kumpulan opini yang masih mengandung *stopword*. Sebagaimana terbukti pada penelitian sebelumnya [21], klasifikasi sentimen teks yang tidak memuat *stopwords* menghasilkan nilai akurasi yang lebih tinggi dibandingkan teks yang memuat *stopwords*. Setelah *stopwords* dihilangkan, dilakukan *stemming* agar setiap kata dihilangkan imbuhanannya sehingga menjadi bentuk kata dasar. Data opini setelah *preprocessing* maka menjadi data bersih yang mudah untuk diklasifikasikan.

### 2.3. Naïve Bayes

Algoritma Naïve Bayes dapat mengklasifikasikan *tweet* berdasarkan polaritasnya, yaitu positif, netral atau negatif [1]. Naïve Bayes dapat diandalkan karena dapat bekerja dengan baik bahkan pada dataset yang kecil [22]. Naïve Bayes adalah metode yang dapat mengklasifikasikan dengan cara menentukan probabilitas kelas tertinggi suatu data yang belum diketahui kelasnya berdasarkan kumpulan data sebelumnya. Walaupun independensi antar fitur tidak terpenuhi sepenuhnya, performa Naïve Bayes tetap mampu mengklasifikasikan dengan baik dan akurat. Sehingga Naïve Bayes banyak dipakai untuk mendukung pengambilan keputusan, prediksi, dan klasifikasi emosi dan sentimen opini [23][18][19].

Algoritma Naïve Bayes dalam penelitian ini berfungsi sebagai pelabelan dan pengujian. Pelabelan kategori sentimen pada opini tidak secara manual, namun dengan *library sentiment* pada fungsi *classify\_polarity* akan mengklasifikasikan opini kedalam kategori sentimen (polaritas) positif, netral, dan negatif dengan metode Naïve Bayes [24]. Pelabelan kategori emosi juga menggunakan metode Naïve Bayes pada *library sentiment* dengan memanggil fungsi *classify\_emotion* yang akan mengklasifikasikan opini kedalam emosi kemarahan (*anger*), kejjikan (*disgust*), ketakutan (*fear*), kegembiraan (*joy*), campuran (*mixed*), kesedihan (*sadness*), dan keterkejutan (*surprise*).

Proses klasifikasi untuk pelabelan kategori sentimen dengan cara menghitung *product* (total perkalian) dari probabilitas setiap *term* (kata) yang muncul pada setiap data opini yang juga terdapat pada leksikon kata bersentimen positif ataupun negatif dalam *library*. Sehingga pada setiap opini akan diketahui probabilitas kata yang bersentimen positif dan probabilitas kata yang bersentimen negatif. Hasil selisih atau agregat dari *product* probabilitas sentimen negatif dan

positif pada opini akan menentukan apakah sentimen opini tersebut positif, negatif, atau netral. Sentimen netral diperoleh jika agregat sangat mendekati atau sama dengan nol.

Pelabelan kategori emosi juga menggunakan proses yang serupa. Hasil dari *product* probabilitas setiap kata dalam opini yang terdapat pada leksikon 6 kategori emosi akan menentukan seberapa probabilitas setiap jenis emosi. Jika probabilitas dari setiap jenis emosi tidak menunjukkan adanya agregat salah satu kategori emosi tertinggi maka opini tersebut mendapatkan label emosi "mixed" atau campuran[19][18]. Untuk pengujian dan evaluasi, dataset opini yang telah mendapatkan label emosi dan label sentimen akan diklasifikasikan menggunakan Naïve Bayes secara 10 *fold cross validation* untuk melihat performa akurasi pada 8 skenario pengujian yakni;

1. Klasifikasi kategori emosi (SA\_emotion) dengan melibatkan fitur teks opini dan fitur polaritas (SA\_polarity)
2. Klasifikasi kategori emosi (SA\_emotion) dengan melibatkan fitur teks opini tanpa fitur polaritas (SA\_polarity)
3. Klasifikasi SA\_polarity dengan melibatkan fitur teks opini dan fitur SA\_emotion
4. Klasifikasi SA\_polarity dengan melibatkan fitur teks opini tanpa fitur SA\_emotion
5. klasifikasi SA\_emotion dengan melibatkan fitur teks opini dan fitur Pol\_sentimentr (polaritas)
6. Klasifikasi SA\_emotion dengan melibatkan fitur teks opini tanpa fitur Pol\_sentimentr
7. Klasifikasi Pol\_sentimentr dengan melibatkan fitur teks opini dan fitur SA emotion
8. klasifikasi Pol\_sentimentr dengan melibatkan fitur teks opini tanpa fitur SA emotion

Fitur SA\_emotion adalah fitur atau atribut berupa label emosi yang diperoleh dari pelabelan dengan Naïve Bayes oleh *library sentiment*. Fitur SA\_polarity adalah label sentimen yang didapatkan dari pelabelan dengan Naïve Bayes oleh *library sentiment*. Sedangkan fitur Pol\_sentimentr adalah fitur atau atribut berupa label sentimen yang diperoleh dari pelabelan menggunakan metode *Valence Shifter* oleh *library sentimentr*. Dengan pengujian dan evaluasi pada 8 skenario diatas akan diketahui bagaimana pengaruh fitur emosi jika dilibatkan dalam klasifikasi sentimen dan sebaliknya. Selain itu akan diketahui pula metode manakah antara Naïve Bayes atau *Valence Shifter* untuk pelabelan sentimen yang menghasilkan model akurasi terbaik. Perhitungan Naïve Bayes untuk menentukan sentimen dapat diilustrasikan pada rumus berikut[25].

$$P(w_k|v_j) = \frac{(n_k+1)}{n + |Vocabulary|} \quad (1)$$

$$P(v_j) = \frac{|docs j|}{|Sampel|} \quad (2)$$

$w_k$  : adalah kata (*word*) k (kata kesatu, kedua, dan seterusnya) dalam semua dokumen (*review*) yang diberi label sebagai j (positif/negatif).

$V_j$  : semua kata (*vocabulary*) dalam kelas j (positif/negatif).

**docs j**: banyaknya dokumen pada kategori j (negatif ataupun positif)

**Sampel**: banyaknya seluruh dokumen ketika tahap training.

$n_k$  : berapa kali suatu kata muncul dalam kelas  $j$ .

$n$  : banyak kata dalam kelas  $j$ .

**Vocabulary** : total kata unik dalam semua dokumen.

$P(w_k|v_j)$  : probabilitas kemunculan kata yang berelasi dengan kelas positif ataupun Negatif.

Nilai dari product probabilitas  $P(w_k|v_j)$  dengan  $P(V_j)$  antara kelas positif dan kelas negatif dibandingkan, sehingga kelas dengan nilai tertinggi akan menjadi label (kelas) bagi data yang akan diklasifikasikan atau data yang belum memiliki kelas (data uji).

#### 2.4. Valence Shifter

Metode *Valence Shifter* dalam penelitian ini juga digunakan untuk pelabelan data opini. *Valence shifter* meliputi, *negator* (misalkan: "never", "none"), *amplifier* (misalkan: "very"), *de-amplifier* (misalkan: "few", "little bit"), *adversative* (misalkan: "but", "instead") dan *contrasting conjunctions* (misalkan: "however"). Setiap kata-kata dari *Valence Shifter* memiliki skor polaritas tertentu yang telah tercakup dalam *leksikon* pada *library*[26] [15]. Metode ini terdapat pada fungsi *sentiment\_by* dalam *library* sentimentr sehingga nilai polaritas dari setiap data komentar *tweet* (opini) diperoleh dengan menjumlah skor setiap kata-kata beserta *Valence Shifter* berdasarkan *leksikon* pada *library*. Jika hasilnya nol maka diklasifikasikan dalam sentimen netral, jika  $> 0$  maka sentimen positif, dan sebaliknya adalah sentimen negatif [15][27]. Polaritas adalah intensitas sentimen yang terkandung dalam teks yang menunjukkan perilaku orang tersebut. Nilai polaritas berkisar dari 1 hingga +1 [28]. Perhitungan skor polaritas menggunakan metode *Valence Shifter* pada suatu opini dapat diilustrasikan pada contoh berikut[26].

"Although(-2) Boris is brilliant(+2) at math, he is a horrible(-2) teacher."

*Valence of term* dasarnya adalah "horrible" (-2) dan "brilliant" (+2), sehingga jika dijumlahkan menjadi nol. Selanjutnya dengan melibatkan skor pada *Valence Shifter* berupa "although" (-2), maka nilai polaritas total dari contoh kalimat tersebut menjadi  $0+ (-2)= (-2)$ .

Nilai skor yang terdapat pada setiap *term* (kata) diatas berdasarkan daftar kata dan skornya dalam *leksikon* pada *library*.

Polaritas (nilai sentimen) pada setiap kalimat ( $r_{sij}$ ) dihitung dengan menjumlahkan semua bobot (skor) dari *term* ( $r_{wijk}$ ) pada suatu kalimat ( $p_j$ ) dibagi dengan akar dari panjang kalimat  $p_j$ . Panjang kalimat ( $p_j$ ) adalah banyaknya *term* pada  $p_j$ [29].

$$r_{sij} = \frac{\sum_{k=1}^{p_j} r_{wijk}}{\sqrt{p_j}} \quad (3)$$

Berdasarkan rumus 3[29], maka nilai polaritas (sentimen) menurut metode *Valence Shifter* pada contoh tersebut adalah negatif karena nilai -2 jika dibagi dengan panjang kalimat yaitu 11, tetap menghasilkan nilai negatif.

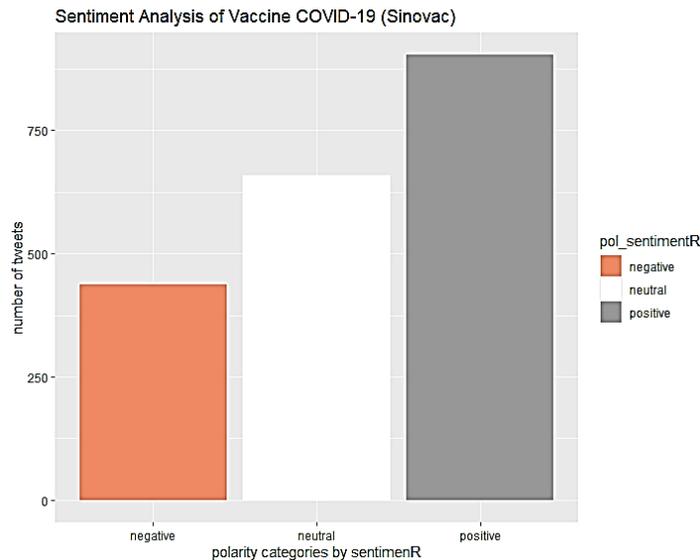
### 3. Hasil dan Pembahasan

Setelah melalui preprocessing maka data menjadi bersih dimana setiap kata telah diubah kedalam bentuk dasarnya (*token*) sebagaimana Gambar 1.

```
"23", "woke peopl realli work china d pimp sinovac just say u  
"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 si  
"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 1. Data Opini setelah *Preprocessing*

Data opini yang telah bersih kemudian dilakukan pelabelan sentimen menggunakan metode *Valence Shifter* yang terdapat pada fungsi *sentiment\_by* dalam *library* *sentimentr*. Proses klasifikasi dengan metode ini dengan cara menghitung skor valensi total dari setiap kata dasar dan kata *Valence Shifter* yang terdapat dalam setiap kalimat opini. Panduan skor setiap kata bernuansa sentimen positif ataupun negatif terdapat dalam leksikon pada *library* *sentimentr*. Hasil pelabelan dengan metode ini sebagaimana Gambar 2 memunculkan 903 (45,15%) sentimen positif, 437 (21,85%) sentimen negatif, dan sentimen netral sebesar 660 (33%).

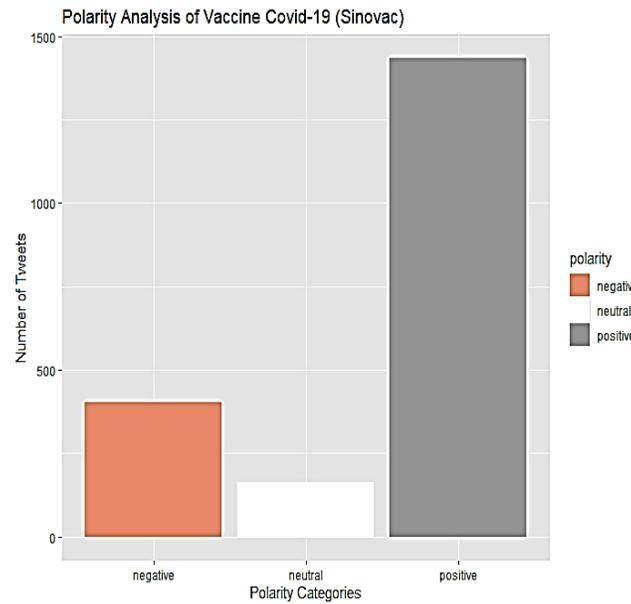


Gambar 2. Hasil Pelabelan dengan Metode *Valence Shifter*

Berdasarkan Gambar 2, Metode *Valence Shifter* memberikan label netral yang cukup banyak yakni 33% dari jumlah total. Hal tersebut sangat mungkin terjadi karena adanya suatu kalimat opini yang sebenarnya bersentimen positif namun karena terdapat kata *Valence Shifter* negatif didalamnya sehingga skor total sentimen dari opini tersebut menjadi netral. Sebagaimana contoh pada metode penelitian diatas, bahwa suatu kalimat yang sebenarnya netral juga dapat menjadi bersentimen negatif jika terdapat *term Valence Shifter* negatif. Dalam metode ini, setiap kata bernuansa sentimen positif, negatif, dan *Valensi Shifter* memiliki skor, maka ini dapat memunculkan bias, karena skor term tersebut terpengaruh dari subjektivitas. Oleh karena itu

dalam klasifikasi yang menggunakan *machine learning*, *Valence Shifter terms* (seperti *although*, *very*, *few*, *never* dan sebagainya) termasuk dalam *stopwords* yang harus dihilangkan.

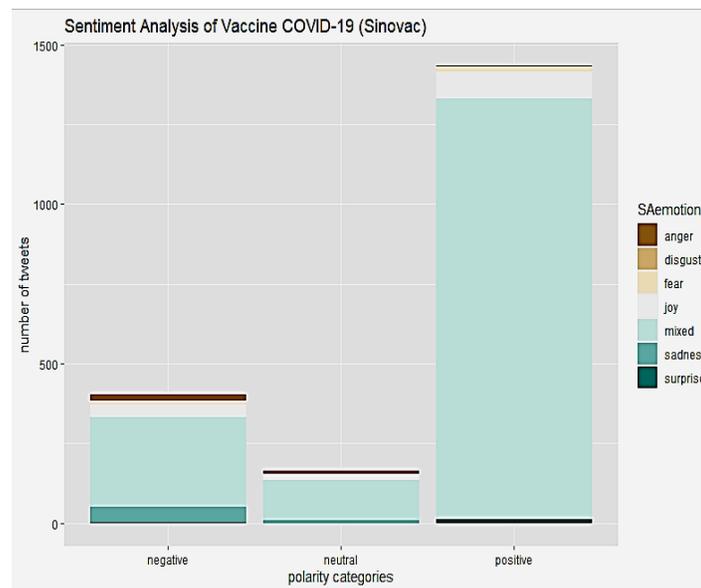
Untuk membandingkan dengan metode *Valence Shifter*, setiap data opini yang telah bersih juga diberi label sentimen dengan cara diklasifikasikan menggunakan metode Naive Bayes dengan memanggil fungsi `classify_polarity` dalam *library sentiment*. Sehingga setiap opini tersebut diberi label positif, negatif, atau netral berdasarkan hasil nilai probabilitas tertinggi dari kata dalam setiap opini yang berelasi dengan jenis sentimen tertentu. Hasil pelabelan sentimen dengan metode Naive Bayes pada *library sentiment* tampak pada Gambar 3.



**Gambar 3.** Hasil Pelabelan Opini dengan Metode Naïve Bayes

Berdasarkan penelitian ini, pelabelan dari klasifikasi dengan metode Naive Bayes mendapatkan sebanyak 1433 (71,65%) sentimen positif, 403 (20,15%) sentimen negatif, dan 164 (8,2%) sentimen netral. Hasil pada Gambar 3 tersebut memiliki diferensiasi yang lebih baik dalam mengenali kelas negatif dan positif dengan ditandai persentase jumlah hasil klasifikasi netral yang lebih kecil daripada hasil pada pelabelan menggunakan metode *Valence Shifter* pada Gambar 2.

Klasifikasi sentimen dan emosi menggunakan *library sentiment* menghasilkan hubungan antara distribusi emosi dan polaritas (kategori sentimen) seperti yang ditunjukkan pada Gambar 4. Dalam hasil ini, setiap kategori sentimen didominasi oleh emosi campuran. Kategori emosi dianggap campuran jika agregat probabilitas antar kategori emosi dalam suatu opini memiliki skor yang sangat kecil atau tidak ada probabilitas emosi yang paling menonjol dari semua kategori emosi pada suatu opini tersebut. Gambar 4 juga menunjukkan bahwa polaritas negatif paling banyak dipengaruhi oleh emosi marah, jijik, gembira, dan sedih. Dalam sentimen negatif, persentase kegembiraan dan kesedihan hampir sama, sehingga kita dapat mengetahui bahwa persentase kemarahan dan jijik sangat mempengaruhi polaritas negatif. Pada polaritas positif, proporsi kegembiraan lebih besar daripada proporsi kegembiraan pada jenis polaritas negatif dan netral. Selain itu, emosi campuran sebagian besar terkandung dalam polaritas positif.

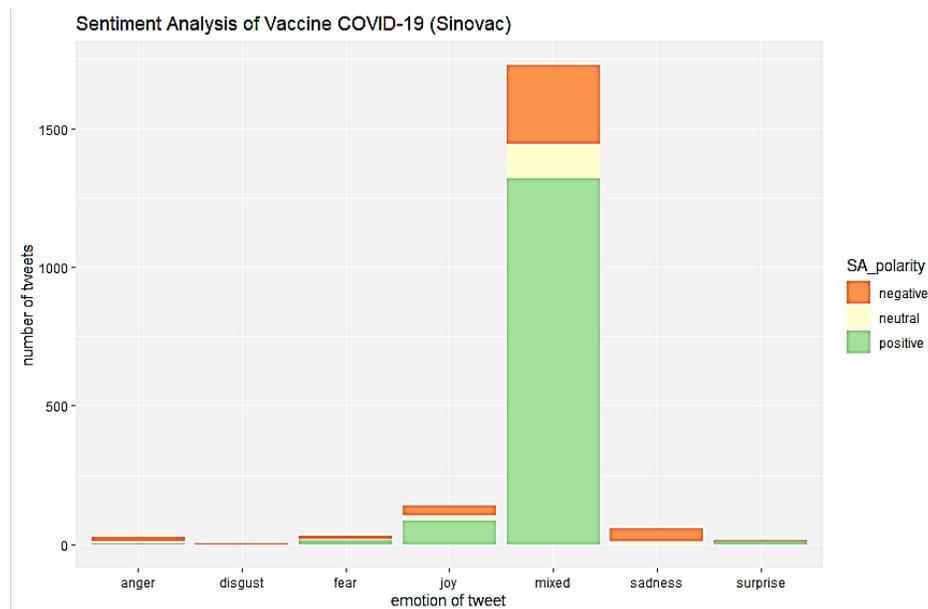


**Gambar 4.** Grafik Intensitas Jenis Emosi Berdasarkan Polaritas Sentimen

Untuk memperjelas hubungan antara emosi dan sentimen, visualisasi data dimunculkan untuk menunjukkan persentase sentimen berdasarkan jenis emosi seperti yang ditunjukkan pada Gambar 4. Pada Gambar 4, dapat terlihat bahwa kategori emosi:

- kemarahan terkait dengan sentimen negatif,
- ketakutan terkait dengan sentimen positif,
- kegembiraan terkait dengan sentimen positif,
- campuran cenderung lebih terkait dengan sentimen positif, negatif dan Netral,
- kesedihan terkait dengan sentimen negatif,
- kejutan cenderung terkait dengan sentimen positif.

Berdasarkan pada Gambar 4, meskipun beberapa kategori emosi terkait dengan sentimen tertentu namun kita melihat bahwa selain kategori emosi campuran tidak terdapat emosi spesifik yang mendominasi. Dengan kata lain, dapat kita asumsikan bahwa tidak ada emosi spesifik yang mencirikan sentimen tertentu, karena emosi campuran mendominasi di semua kategori sentimen. Sehingga berdasarkan data *tweet* yang diperoleh, sangat mungkin label emosi kurang signifikan untuk membantu menentukan kategori sentimen. Namun, terdapat kemungkinan bahwa label sentimen justru signifikan dalam mencirikan atau menentukan kategori emosi. Untuk memperjelas asumsi ini, visualisasi berikut ditampilkan pada Gambar 5 untuk melihat distribusi kategori sentimen pada setiap kategori emosi.



**Gambar 5.** Intensitas Kategori Sentimen pada setiap Kategori Emosi

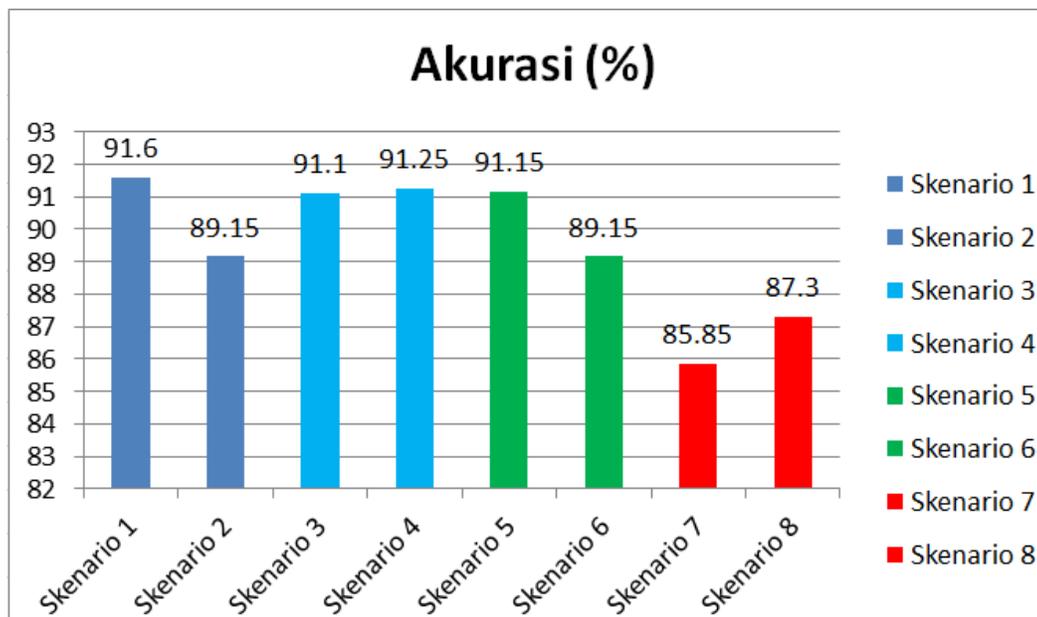
Gambar 5 memperjelas asumsi sebelumnya, bahwa hampir di setiap kategori emosi terdapat kategori sentimen tertentu yang mendominasi atau mencirikan. Pada emosi *mixed* atau campuran, didominasi atau dicirikan oleh sentimen positif. Pada emosi *sadness* didominasi oleh sentimen negatif. Pada emosi kegembiraan atau *joy* dicirikan oleh sentimen positif. Pada emosi *anger* didominasi oleh sentimen negatif. Sehingga semakin jelas menurut Gambar 5 bahwa kategori emosi sangat dipengaruhi oleh kategori sentimen, atau sentimen menentukan kategori emosi.

Pengujian dan evaluasi perlu dilakukan sebagai tahap berikutnya untuk melihat apakah benar kategori polaritas (sentimen) dapat menentukan kategori emosi, atau apakah benar bahwa fitur sentimen dapat menjadi fitur yang signifikan meningkatkan akurasi untuk menentukan kategori emosi. Pada tahap ini data opini memiliki fitur: teks, kategori sentimen (SA\_Polarity), kategori emosi (SA\_emotion) dan Kategori Sentimen (Pol\_sentimentr). Fitur SA\_emotion adalah fitur atau atribut berupa label emosi yang diperoleh dari pelabelan dengan Naïve Bayes oleh *library sentiment*. Fitur SA\_polarity adalah label sentimen yang didapatkan dari pelabelan dengan Naïve Bayes oleh *library sentiment*. Sedangkan fitur Pol\_sentimentr adalah fitur atau atribut berupa label sentimen yang diperoleh dari pelabelan menggunakan metode *Valence Shifter* oleh *library sentimentr*. Pengujian dilakukan dengan klasifikasi menggunakan Naïve Bayes secara *10 fold cross validation* pada 8 skenario.

- 1) Klasifikasi kategori emosi (SA\_emotion) dengan melibatkan fitur teks opini dan fitur polaritas (SA\_polarity)
- 2) Klasifikasi kategori emosi (SA\_emotion) dengan melibatkan fitur teks opini tanpa fitur polaritas (SA\_polarity)
- 3) Klasifikasi SA\_polarity dengan melibatkan fitur teks opini dan fitur SA\_emotion
- 4) Klasifikasi SA\_polarity dengan melibatkan fitur teks opini tanpa fitur SA\_emotion
- 5) klasifikasi SA\_emotion dengan melibatkan fitur teks opini dan fitur Pol\_sentimentr (polaritas)

- 6) Klasifikasi SA\_emotion dengan melibatkan fitur teks opini tanpa fitur Pol\_sentimentr
- 7) Klasifikasi Pol\_sentimentr dengan melibatkan fitur teks opini dan fitur SA emotion
- 8) klasifikasi Pol\_sentimentr dengan melibatkan fitur teks opini tanpa fitur SA emotion

Hasil akurasi dari setiap skenario pengujian ditampilkan dalam Gambar 6.



**Gambar 6.** Hasil Pengujian dari 8 Skenario

Berdasarkan Gambar 6 dan Tabel 1, pada skenario 1 dan skenario 5 menunjukkan nilai akurasi yang meningkat ketika menentukan kelas emosi dengan melibatkan fitur sentimen (polaritas). Pada skenario 2 dan skenario 6, nilai akurasi klasifikasi emosi justru menurun ketika tidak melibatkan fitur polaritas (sentimen).

**Tabel 1.** Hasil Pengujian pada 8 Skenario

Skenario Pengujian	Kelas (Label)	Fitur yang dilibatkan	Akurasi (%)
Skenario 1	SA_emotion	teks ; SA_polarity	91.6
Skenario 2	SA_emotion	teks	89.15
Skenario 3	SA_polarity	teks; SA_emotion	91.1
Skenario 4	SA_polarity	teks	91.25
Skenario 5	SA_emotion	teks;Pol_sentimentr	91.15
Skenario 6	SA_emotion	teks	89.15
Skenario 7	Pol_sentimentr	teks; SA_emotion	85.85
Skenario 8	Pol_sentimentr	teks	87.3

Hasil skenario 3 dan 4 menunjukkan bahwa nilai akurasi justru menurun ketika mengklasifikasikan sentimen dengan melibatkan fitur kategori emosi. Begitu juga pada hasil 7 dan 8, nilai akurasi meningkat ketika mengklasifikasikan sentimen tanpa melibatkan fitur kategori emosi. Hasil ini menunjukkan berdasarkan data *tweet* yang digunakan dalam penelitian ini, fitur emosi (SA\_emotion) kurang baik dalam menentukan sentimen (SA\_polarity maupun Pol\_sentimentr), namun fitur sentimen (SA\_polarity maupun Pol\_sentimentr) cukup baik dalam mencirikan atau menentukan emosi (SA\_emotion).

#### 4. Kesimpulan dan Saran

Pelabelan sentimen dengan metode Naive bayes maupun *Valence Shifter* menghasilkan persentase sentimen positif terbanyak dari data *tweet* berkata kunci “sinovac”. Metode Naive Bayes lebih objektif dalam menentukan label sentimen karena menggunakan agregat probabilitas *term* dengan jenis sentimen yang muncul pada seluruh dokumen teks dan tidak memperhatikan kata-kata *Valence Shifter*. Sedangkan metode *Valence Shifter* menggunakan skor setiap kata dasar dan kata *Valence shifter* yang terdapat dalam kalimat teks sehingga lebih rentan terjadi bias sentimen. Berdasarkan hasil penelitian ini, hubungan antara jenis sentimen dan kategori emosi menunjukkan bahwa fitur sentimen dapat menjadi prediktor yang baik untuk menentukan kategori emosi. Sebaliknya, fitur emosi kurang baik jika menjadi prediktor untuk menentukan sentimen. Untuk penelitian kedepannya, akan lebih baik jika digunakan data yang lebih banyak, dengan metode klasifikasi yang berbeda serta data yang digunakan selain berbahasa Inggris, misalkan data *tweet* berbahasa Indonesia. Disamping itu, kategori emosi campuran mungkin dapat diklasifikasikan dengan data *tweet* dengan kategori emosi tertentu sehingga akan mengurangi persentase kategori emosi campuran.

## 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] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, p. 112, 2020, doi: 10.20473/jisebi.6.2.112-122.
- [4] "During The Pandemic, Twitter Users Grew Rapidly." <https://voi.id/en/technology/48500/during-the-pandemic-twitter-users-grew-rapidly> (accessed Sep. 22, 2021).
- [5] X. Guo and J. Li, "A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency," in *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, 2019, no. 1, pp. 472–477, doi: 10.1109/SNAMS.2019.8931720.
- [6] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*, 2010, pp. 627–666.
- [7] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017, doi: 10.22146/ijccs.24716.
- [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] T. Danisman and A. Alpkocak, "Feeler: Emotion classification of text using vector space model," in *AISB 2008 Convention: Communication, Interaction and Social Intelligence - Proceedings of the AISB 2008 Symposium on Affective Language in Human and Machine*, 2008, vol. 1, pp. 53–59.
- [10] S. An, L. J. Ji, M. Marks, and Z. Zhang, "Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures," *Front. Psychol.*, vol. 8, no. APR, pp. 1–14, 2017, doi: 10.3389/fpsyg.2017.00610.
- [11] A. Sharma and U. Ghose, "Sentimental Analysis of Twitter Data with respect to General Elections in India," in *Procedia Computer Science*, 2020, vol. 173, no. 2019, pp. 325–334, doi: 10.1016/j.procs.2020.06.038.
- [12] H. Deng *et al.*, "Sentiment analysis of real-world migraine tweets for population research," *Cephalalgia Reports*, vol. 3, pp. 1–9, 2020, doi: 10.1177/2515816319898867.

- [13] S. S. and P. K.V., “Sentiment analysis of malayalam tweets using machine learning techniques,” *ICT Express*, vol. 6, no. 4, pp. 300–305, 2020, doi: 10.1016/j.icte.2020.04.003.
- [14] K. Mohamed Ridhwan and C. A. Hargreaves, “Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100021, 2021, doi: 10.1016/j.jjime.2021.100021.
- [15] 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>.
- [16] 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>.
- [17] 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>.
- [18] B. M. Akbar, A. T. Akbar, and R. Husaini, “Classification of Sentiments on Twitter Opinions with The Keyword Sinovac Using Naive 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>.
- [19] B. M. Akbar, A. T. Akbar, and R. Husaini, “Analisis Sentimen dan Emosi Vaksin Sinovac pada Twitter menggunakan Naïve Bayes dan Valence Shifter,” *J. Teknol. Terpadu*, vol. 7, no. 2, pp. 83–92, 2021, doi: 10.54914/jtt.v7i2.433.
- [20] 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.
- [21] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016.
- [22] RapidMiner, “Naive Bayes - RapidMiner Documentation,” *RapidMiner*, 2019. [https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive\\_bayes.html](https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive_bayes.html) (accessed Sep. 22, 2021).
- [23] I. Rish, “An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive 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](https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier/link/00b7d52dc3ccd8d692000000/download).
- [24] E. Riloff and J. Wiebe, “Learning extraction patterns for subjective expressions,” pp. 105–112, 2003, doi: 10.3115/1119355.1119369.
- [25] F. Handayani and S. Pribadi, “Implementasi Algoritma Naive Bayes Classifier dalam

- Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110,” *J. Tek. Elektro*, vol. 7, no. 1, pp. 19–24, 2015.
- [26] L. Polanyi and A. Zaenen, “Contextual valence shifters,” *AAAI Spring Symp. - Tech. Rep.*, vol. SS-04-07, no. August, pp. 106–111, 2005, doi: 10.1007/1-4020-4102-0\_1.
- [27] M. Naldi, “A review of sentiment computation methods with R packages,” no. January, 2019, [Online]. Available: <http://arxiv.org/abs/1901.08319>.
- [28] M. Singh, A. K. Jakhar, and S. Pandey, “Sentiment analysis on the impact of coronavirus in social life using the BERT model,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1007/s13278-021-00737-z.
- [29] S. Balbi, M. Misuraca, and G. Scepi, “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.