



Analisis Sentimen Mengenai Vaksin *Sinovac* di Media Sosial Twitter Menggunakan Metode *Naïve bayes Classification*

Rima Tamara Aldisa ^{1*}, Azizah ², Mohammad Aldinugroho Abdullah ³

^{1,2,3} Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

article info

Article history:

Received 16 September 2021

Received in revised form

3 October 2021

Accepted 5 October 2021

Available *online* July 2022

DOI:

<https://doi.org/10.35870/jtik.v6i3.479>

Keywords:

Sentiment Analysis; Naïve bayes classification; Twitter; Sinovac.

Kata Kunci:

Analisis Sentimen; Klasifikasi Naïve bayes; Twitter; Sinovac.

abstract

The Sinovac vaccine is an example of a type of inactivated vaccine. The government bought Sinovac, Novavax, AstraZeneca, and Pfizer vaccines. This vaccine is used to treat the Covid-19 pandemic. This vaccine is used to treat the Covid-19 pandemic. The role of the Indonesian people in expressing and stating the pros and cons often involves public services that are easily accessible by many people, namely social media, one of which is Twitter. This can be used as material to analyze who produces data in support of decisions. The technique that can be used is sentiment analysis. The method used in this study is the Naïve bayes Classification. The purpose of this study was to determine the value of sentiment analysis on the Sinovac vaccine using the Naïve Bayes Classification method on Twitter social media using Indonesian. The result of this research is the final probability value based on the condition 0.000002765 for positive and 0.000000359 for negative. A response with a positive comment has a greater probability of a response with a negative comment.

abstrak

Vaksin Sinovac adalah contoh dari jenis vaksin inaktivasi. Pemerintah membeli vaksin Sinovac, Novavax, AstraZeneca, dan Pfizer. Vaksin ini digunakan untuk menangani pandemi Covid-19. Vaksin ini digunakan untuk menangani pandemi Covid-19. Peran masyarakat Indonesia dalam mengungkapkan dan menyatakan pro dan kontra seringkali melibatkan layanan publik yang mudah diakses oleh banyak orang, yaitu media sosial salah satunya adalah Twitter. Hal ini dapat digunakan sebagai bahan untuk menganalisis siapa yang menghasilkan data dalam mendukung keputusan. Teknik yang dapat digunakan adalah analisis sentimen. Metode yang digunakan dalam penelitian ini adalah Klasifikasi Naïve bayes. Tujuan dari penelitian ini adalah untuk mengetahui nilai analisis sentimen pada vaksin Sinovac dengan menggunakan metode Naïve bayes Classification pada media sosial Twitter menggunakan Bahasa Indonesia. Hasil dari penelitian ini adalah nilai probabilitas akhir berdasarkan kondisi 0,000002765 untuk positif dan 0,000000359 untuk negatif. Sebuah respon dengan komentar positif memiliki probabilitas lebih besar kemungkinan respon dengan komentar negatif.

Corresponding author. Email: rimatamaraaldisa@gmail.com ^{1}.

1. Latar Belakang

Indonesia sedang dilanda pandemi Covid-19 dimana kasus pertama dilaporkan di Indonesia pada 2 Maret 2020 dengan total dua kasus [1]. Pemerintah Indonesia di lingkungan pandemi Covid-19 menerapkan Pembatasan Sosial Berskala Besar (PSBB) dimana masyarakat diharuskan melakukan aktivitas seperti bekerja dan belajar dari rumah. Dengan adanya pandemi ini berdampak pada setiap sektor pemerintahan Indonesia. Oleh karena itu, pemerintah Indonesia melakukan upaya vaksinasi di masyarakat untuk mengatasi masalah Covid-19. Vaksin Sinovac adalah contoh dari jenis vaksin yang tidak aktif. Pemerintah telah membeli vaksin untuk Sinovac, Novavax, AstraZeneca, dan Pfizer [2].

Vaksin digunakan untuk mengatasi dengan pandemi Covid-19. Di lingkungan pandemi ini, peran serta masyarakat juga diperlukan dengan mematuhi protokol kesehatan dan pembongkaran fisik agar tidak tertular virus Covid-19. Selain itu, peran masyarakat juga diperlukan dalam menerima dan mendukung program vaksinasi Covid-19. Salah satu cara pemerintah dalam membentuk kepercayaan masyarakat, dengan melakukan vaksinasi Presiden Republik Indonesia Joko Widodo yang disiarkan secara langsung dan dapat dilihat oleh masyarakat Indonesia. Namun, hal tersebut tidak cukup untuk membangun kepercayaan publik dan menimbulkan pro dan kontra di masyarakat Indonesia. Masyarakat Indonesia dalam mengungkapkan dan menyatakan pro dan kontra seringkali melibatkan layanan publik yang mudah diakses oleh banyak orang yaitu media sosial salah satunya Twitter. Twitter adalah situs populer yang digunakan dalam mengunggah tweet dengan batas karakter 140 karakter [3].

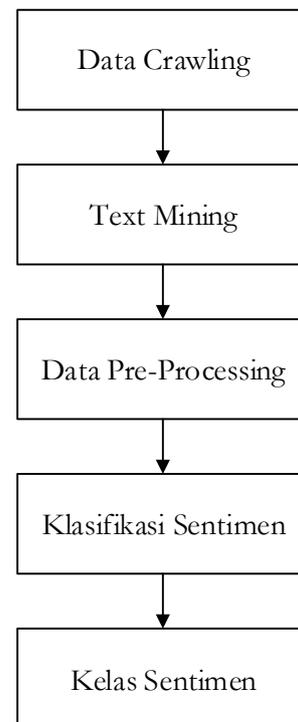
Dalam melakukan analisis sentimen ini terdapat penelitian yang telah dilakukan sebelumnya yang menjadi bahan referensi bagi peneliti. Penelitian yang dilakukan oleh Astari dkk pada tahun 2020 melakukan penelitian dengan menggunakan metode *Naïve bayes Classifier* yang menghasilkan *tweet* dengan dampak virus corona dengan hasil yang stabil [4]. Selanjutnya penelitian yang dilakukan oleh Nugroho pada tahun 2018 menggunakan metode yang sama yaitu *Naïve bayes Classifier* dengan hasil yang diperoleh yaitu nilai negatif dari topik yang diangkat oleh peneliti [3].

Kemudian penelitian yang dilakukan oleh Ikasari dkk pada tahun 2020 melakukan metode klasifikasi dengan hasil analisis sentimen positif dari topik yang diangkat oleh peneliti [5].

Berdasarkan latar belakang dalam penelitian ini maka penulis melakukan penelitian ini dengan tujuan untuk mengetahui nilai analisis sentimen pada vaksin Sinovac menggunakan metode *Naïve bayes Classification* pada media sosial Twitter menggunakan bahasa Indonesia. rumusan masalah penelitian ini adalah perlunya melakukan analisis sentimen mengenai vaksin sinovac di media sosial twitter menggunakan metode *naïve bayes classification*. Tujuan penelitian ini adalah untuk membangun kepercayaan publik dan mengurangi adanya pro dan kontra di masyarakat Indonesia mengenai vaksin Sinovac.

2. Metode Penelitian

Alur kerja Analisis Sentimen yang digunakan dalam penelitian ini dapat diilustrasikan pada Gambar 1. Dimulai dengan melakukan data *crawling* atau pengumpulan data dari media sosial Twitter. Kemudian lakukan *text mining* yang meliputi *tokenization*, *case folding*, dan *stemming*. Dan lanjutkan dengan mengklarifikasi sentimen menggunakan *Naïve Bayes*.



Gambar 1. *Sentiment Analysis Workflow*

Instrumen Penelitian

Dalam melakukan penelitian ini, menggunakan bahasa pemrograman Python dengan Tweepy sebagai library utama. Kemudian, sumber data yang telah diperoleh pada data *crawling* diolah menggunakan Microsoft Excel terkait dengan proses pra-pemrosesan dan analisis.

Prosedur Pengumpulan Data

Teknik pengumpulan data yang dilakukan dalam penelitian ini menggunakan bahasa pemrograman Python. Pengamatan pada media sosial Twitter dilakukan dengan menggunakan teknik *text mining* selama 6 jam pada setiap kata kunci.

Metode Analisis Data

Metodologi proses terdiri dari data *crawling* yang diambil dari media sosial twitter, *text mining*, dan *Naive Bayes Classification*. Python digunakan untuk mengumpulkan data dari API yang digunakan yaitu Twitter API.

1) Perayapan Data

Data *crawling* merupakan salah satu teknik pengumpulan data yang digunakan pada media internet. Proses Data yang dikumpulkan adalah data pengguna media sosial Twitter berbahasa Indonesia dari seluruh pengguna di Indonesia dengan kata kunci “*Sinovac Vaccine*”. Data ini diperoleh dengan sebuah fungsi dalam program Python yang terhubung dengan API (*Application Programming Interface*) milik Twitter.

2) Penambahan Teks

Text mining adalah mencari data yang akurat dalam dokumen teks dalam mencari data yang diinginkan [6]. Data yang telah diperoleh akan melakukan tahap tokenization dimana kalimat dipecah menjadi kata-kata yang berurutan (token). Data berupa token memasuki tahap case folding dimana huruf besar pada data akan diubah menjadi huruf kecil. Data kemudian masuk ke tahap stemming dimana data hanya akan diambil dari kata dasar dan dibuang imbuhanannya.

3) Prapemrosesan Data

Preprocessing data adalah teknik manipulasi data sebelum data diklasifikasikan. Dalam *preprocessing* data, data penting akan dipisahkan dari data yang tidak penting dan akan terjadi proses menghilangkan data mentah yang

telah diperoleh sebelumnya.

4) Klasifikasi

Pada tahap klasifikasi ini kami menggunakan metode klasifikasi *Naive Bayes*.

3. Hasil Analisis dan Desain

Analisis

Tahap pertama dalam penelitian ini adalah data *crawling*. Proses *crawling* data dilakukan dengan menggunakan bahasa pemrograman Python selama 8 jam. Berikut ini adalah algoritma yang digunakan dalam program data *crawling*

Program Data Crawling

```
class StdOutListener (StreamListener):
    def on_status(self, status):
        try:
            with open('vaksin.txt', 'a')
            as record:
                record.write(str(status.text) + '\n')
            return True
        except BaseException as e:
            print("Error on_data: %s" % str(e))
            print(status.text)
            return True
    def on_error(self, status_code):
        if status_code == 420:
            return False
if name == 'main':
```

Data yang diperoleh pada proses ini berupa *tweet* yang berisi kata kunci terkait dengan vaksin Sinovac. Klasifikasi jenis *tweet* dibagi menjadi 2 jenis, yaitu kata positif dan negatif. Tabel di bawah ini adalah klasifikasi kata berdasarkan klasifikasi jenis kata positif dan negatif.

Tabel 1. Klasifikasi Jenis Kata Positif dan Negatif

Positif	Negatif
Optimistic	danger
Guaranteed	Conspiracy
Safe	Bad
Effective	Hoax
Original	Fake
Halal	Fret

Setelah data *tweet* terkait tanggapan terkait vaksin Sinovac di Twitter terkumpul melalui teknik data *crawling*, selanjutnya akan dilakukan tahap *text mining*. Algoritma yang digunakan pada tahap *text mining* adalah sebagai berikut.

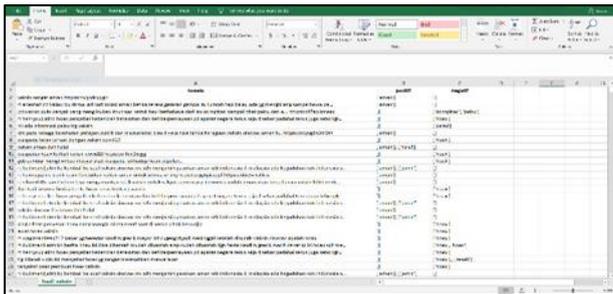
Penambahan Teks Program

```
for tweet in tweets_list:
    text_positif=[]
    text_negatif=[]
    positive_counter = 0
    negative_counter = 0
```

```

tweet_proces_2 = tweet.replace ('!',
''.replace('.', ''))
tweet_processed =
tweet_proces_2.lower()
for p in list (punctuation): tweet_processed
=
tweet_processed.replace(p, '')
clean_tweets.append(tweet_processed)
words = tweet_processed.split(' ') word_count =
len(words)
for word in words:
if word in positive_words:
text_positif.append(word) positive_counter =
positive_counter+1
elif word in negative_words:
text_negatif.append(word) negative_counter =
negative_counter+1
positive_counts.append((positive_counter/
word_count)*100)
negative_counts.append((negative_counter/
word_count)*100)
positive_word.append(text_positif)
negative_word.append(text_negatif)
    
```

Pada tahap ini dilakukan *filter* kata pada *tokenization*, *case folding*, dan *stemming tweet*. Berikut adalah hasil pada tahap text mining:



Gambar 2. Hasil Proses Text Mining.

Gambar 2 menunjukkan hasil yang diperoleh pada tahap text mining. Hasil ini akan segera diklasifikasikan melalui tahap preprocessing berdasarkan probabilitas dan kelas-kelas, yang di kelas sesuai dengan jenis kata positif dan negatif. Hasil dari tahap preprocessing akan diolah melalui proses klasifikasi menggunakan metode Naïve Bayes. Data yang akan diklasifikasikan berupa kumpulan kata-kata positif dan negatif yang terdapat dalam data tweet. Persamaan Naïve Bayes yang akan digunakan pada tahap klasifikasi adalah sebagai berikut.

$$P(c \setminus x) = P(x1 \setminus c) P(x2 \setminus c) P(xn \setminus c) \dots P(c)$$

Keterangan:

- P (c \ x) = Probabilitas dalam kondisi.
- P (xn \ c) = probabilitas menurut kondisi hipotetis.
- P (c) = Probabilitas seluruh hipotesis.

Berikut adalah data yang telah diolah melalui tahapan text mining dan preprocessing yang dapat dilihat pada tabel 2.

Tabel 2. Sentimen Vaksin Sinovac

Word	Type	Amount
Optimistic	Positif	51
Guaranteed	Positif	19
Safe	Positif	41
Effective	Positif	10
Original	Positif	5
Halal	Positif	37
Danger	Negatif	9
Conspiracy	Negatif	8
Bad	Negatif	4
Hoax	Negatif	45
False	Negatif	2
Fret	Negatif	9
Total		240

Tabel 2 menunjukkan data yang telah diklasifikasikan menurut jenisnya, yaitu positif atau negatif. Jumlah kata yang terdapat pada data di atas adalah 240 kata. Untuk kata positif ada 163 kata, dan untuk kata negatif ada 77 kata. Berikut adalah tabel klasifikasi menggunakan metode Naïve Bayes.

Tabel 3. Hasil Klasifikasi Sentimen Responden Vaksin Sinovac

Probabilitas	Positif	Negatif
P(c)	0.705627706	0.320833333
P(x1 \ c)	0.312883436	0.116883117
P(x2 \ c)	0.116564417	0.103896104
P(x3 \ c)	0.251533742	0.051948052
P(x4 \ c)	0.061349693	0.584415584
P(x5 \ c)	0.030674847	0.025974026
P(x6 \ c)	0.226993865	0.116883117
P(c \ x)	0.000002765	0.000000359

Pada tabel 3, terdapat P (c) yang mewakili probabilitas hipotesis keseluruhan, dan P (xn \ c) mewakili probabilitas setiap kelas kata. P(c \ x) diperoleh sebagai nilai probabilitas akhir berdasarkan kondisi melalui perhitungan Naïve Bayes dengan nilai positif sebesar 0,000002765 dan nilai negatif sebesar 0,000000359. Dalam hal ini, respon terhadap topik yang diteliti menuai respon positif yang lebih besar daripada respon negatif.

Dari pengamatan dan analisis yang peneliti lakukan, peneliti ingin membuktikan bahwa vaksin Sinovac mendapat pandangan dan dianggap baik atau buruk oleh orang-orang yang menggunakan Twitter.

4. Kesimpulan

Berdasarkan data tweet terkait tanggapan terkait vaksin Covid-19 di Twitter yang telah dikumpulkan melalui data crawling dan diolah melalui tahapan text mining dan preprocessing, serta diklasifikasikan melalui metode Naïve Bayes, maka nilai probabilitas akhir yang dihasilkan adalah 0,000002765 untuk positif dan 0,000000359 untuk negatif. Dapat disimpulkan bahwa tanggapan dengan komentar positif memiliki probabilitas yang lebih besar daripada kemungkinan tanggapan dengan komentar negatif.

5. Daftar Pustaka

- [1] Susilo, A., C. martin R., dkk. 2020. Coronavirus Disease 2019: Tinjauan Literatur Terbaru. Jurnal Ilmu Penyakit Dalam Indonesia, Vol. 7, No. 1.
- [2] Kementerian Kesehatan Republik Indonesia. 2021. Faktor Kunci Triple-helix Menuju Keberhasilan Vaksinasi Covid-19. Infodatin Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia, ISSN 2442-7659.
- [3] Nugroho, Agung. 2018. Analisis Sentimen pada Media Sosial Twitter Menggunakan Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram. Jurnal Ilmu Komputer & Informatika (J-SAKTI), Jilid (2) No.2.
- [4] Astari, Ni Made A J., Dewa G H D., & Gede I. 2020. Analisis Sentimen Dokumen Twitter Terkait Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. Jurnal Sistem dan Informatika, Vol 15, No 1.
- [5] Ikasari, D., Yuliana F., & Widiastuti. 2020. Analisis Sentimen dan Klasifikasi Tweet Bahasa Indonesia Pada Angkutan Umum Mrt Jakarta Menggunakan Naive Bayes Classifier. Jurnal Ilmiah Informatika Komputer Jilid 25 No. 1.
- [6] Taufik, Andi. 2018. Perbandingan Algoritma Text Mining Untuk Klasifikasi Review Hotel. Jurnal Teknik Komputer AMIK BSI, Vol. IV No.2.