



Analisis Sentimen Di Media Sosial Twitter Dengan Studi Kasus Vaksinasi Covid-19

Nufia Alfi Rohyana^{a,1,*}, Aris Wahyu Murdiyanto^{b,2}, Kharisma^{b,3}

^a Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

^b Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

^b Universitas Jenderal Achmad Yani Yogyakarta, Jalan Siliwangi - Gamping, Sleman 55294, Indonesia

¹ nufiaalfi27@gmail.com*; ² ariswahyu@unjaya.ac.id; ³ kharisma.ano@gmail.com

* corresponding author

ABSTRACT

ARTICLE INFO

With the COVID-19 pandemic, the World Health Organization or WHO conducted research and research trials on the COVID-19 vaccine. The Indonesian government has made several policies, one of which is the "Mass Vaccination Program". However, the COVID-19 vaccination program in the field received mixed responses in the community, there were those who supported the vaccine program and some who rejected the vaccine program. In this study, researchers conducted research on sentiment analysis on the opinion of vaccination programs against anti-vaccine community groups based on Twitter social media data using the Naïve Bayes Classifier algorithm to provide information on opinion assessments that lead to positive and negative sentiments.

Objective: The purpose of this study is to find out the public perception of AntiVaccine against the COVID-19 Vaccination Program in Indonesia.

This study uses the Naïve Bayes Classification. The use of the Naïve Bayes Classifier (NBC).

This research uses tweets obtained from Twitter with the keywords/hashtags "Anti Covid-19 Vaccines" or by collecting data based on accounts related to news about vaccination programs such as @ The Ministry of Health of the Republic of Indonesia. Data collection was carried out in the period August 2021-December 2021, with a total of 889 data. This study has succeeded in obtaining an accuracy of 72 % for testing. The result of the final sentiment analysis in the classification of the Anti-Vaccine group in this study is "Negative Sentimen".

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article history

Received: 21 Februari 2023

Revised: 10 Maret 2023

Accepted: 10 Mei 2023

Keywords

Sentiment analysis

Naïve bayes classifier

Anti covid-19 vaccines

1. Pendahuluan

Pada bulan Desember 2019, dunia dihebohkan dengan kejadian pandemik yang disebabkan oleh Corona Virus/COVID-19. Virus COVID-19 pertama kali terdeteksi di China, kemudian menyebar ke seluruh dunia termasuk juga Negara Indonesia. Virus COVID-19 melanda Indonesia pada awal Maret 2020 [1]. Pada bulan Agustus tahun 2022, menurut laporan Kementerian Kesehatan Republik Indonesia total kasus COVID-19 di Indonesia mencapai 6,26 Juta jiwa, dengan laju kasus baru



sebesar 6,276 kasus setiap harinya [2]. Dengan adanya pandemik COVID-19 tersebut, Organisasi Kesehatan Dunia atau WHO melakukan penelitian dan uji riset mengenai vaksin COVID-19 [3]. Dalam mencegah penyebaran kasus COVID-19 semakin meluas, Pemerintah Indonesia membuat beberapa kebijakan antara lain “New Normal”, “Pembatasan Skala Berskala Besar”, “Work Form Home” serta yang terakhir “Program Vaksinasi Massal” [4]. Pada Program Vaksinasi Massal ini diharapkan dapat memutus rantai pemberantasan Virus COVID-19 secara efektif dan efisien agar tercapainya Herd Immunity untuk mengendalikan laju penyebaran COVID-19 [5]. Vaksin Covid-19 pertama kali tiba di Indonesia pada tanggal 6 Desember 2020 sebanyak 1,2 juta dosis dengan merk CoronaVac buatan dari Sinovac [6]. Setelah berlangsungnya vaksinasi terhadap Presiden Republik Indonesia, program vaksinasi akan dilakukan secara serentak terhadap tenaga kesehatan kemudian masyarakat di 34 Provinsi di Indonesia [7].

Perbedaan yang terjadi di masyarakat muncul akibat adanya disinformasi yang diterima oleh masyarakat berupa informasi mengenai Vaksin Covid-19 [8]. Dari laporan “Digital Indonesia” pengguna media sosial di Indonesia mengalami pertumbuhan yang sangat pesat. Data-data dari media sosial dapat dilakukan analisis untuk mendapatkan wawasan, trending topik, influencer serta informasi lainnya [9]. Cara untuk melakukan analisis terhadap opini yang ada di media sosial salah satunya adalah dengan cara analisis sentimen, analisis sentimen berfokus pada opini positif dan negatif [10]. Pada algoritma Naïve Bayes Classifier (NBC) sering digunakan karena memiliki kelebihan akurasi yang tinggi dan membutuhkan sample data dalam jumlah kecil untuk memperkirakan parameter yang diperlukan dalam klasifikasi [11]. Berdasarkan permasalahan di latar belakang, jumlah cuitan terkait program vaksin pada kelompok Antivaksin Covid di Twitter belum dianalisis secara mendalam, sehingga belum diketahui isi informasi yang sesuai dengan kondisi vaksinasi yang sebenarnya di lapangan.

2. The Proposed Method/Algorithm

2.1. Keaslian Penelitian

Perbedaan penelitian peneliti dibanding dengan peneliti terdahulu [12][13][14][15] dengan topic “Analisis Sentimen pada Media Sosial Twitter” berupa subjek penelitian, pengambilan data serta analisis data tweet. Pada penelitian terdahulu menggunakan tools RStudio dan RapidMiner. Sedangkan pada penelitian sentimen analisis pada Vaksinasi Covid-19 menggunakan program Python dengan tools Jupyter Notebook serta peneliti menggunakan Algoritma NBC dalam klasifikasi.

2.2. Twitter

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunaannya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet). Twitter menyediakan API (Application Programming Interface) diperuntukkan bagi pengembang yang memungkinkan pengguna dapat membaca, menulis dan mengambil data dari Twitter. Penggunaan Twitter API ini juga memungkinkan pengembang untuk mengambil informasi atau data pengguna di Twitter atau suatu subjek di lokasi tertentu [16].

2.3. Text Mining

Text Mining, adalah proses penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru dalam bentuk sebuah informasi yang tersirat secara implisit, informasi yang diekstrak secara otomatis dari sumber sumber data teks yang berbeda-beda. Text mining mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti Data mining, Information Retrieval, Statistik dan Matematik, Machine Learning, Linguistic, Natural Language Processing, dan Visualization. Kegiatan riset untuk text mining antara lain ekstraksi dan penyimpanan text, preprocessing akan konten text, pengumpulan data statistik dan indexing dan analisa konten [17].

2.4. Analisis Sentimen

Analisis sentimen adalah aplikasi text mining yang banyak digunakan yang dapat melacak sentimen pelanggan tentang perusahaan. Juga dikenal sebagai penambangan opini, analisis sentimen menambang teks dari tinjauan online, jejaring sosial, email, interaksi pusat panggilan, dan sumber data lainnya untuk mengidentifikasi untaian umum yang mengarah ke perasaan positif atau negatif dari pihak pelanggan. Analisis sentimen merupakan bentuk teknologi baru yang saat ini sedang

dikembangkan dan digunakan dalam melakukan sebuah penelitian guna menganalisis topik dan juga pendapat [18].

3. Method

Naïve Bayes Classification, adalah metode pengklasifikasian probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi atau kejadian. Naive Bayes Classifier memiliki bentuk umum seperti berikut [19] :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

X	= Data dengan kelas tidak dikenal
H	= Hipotesis data X adalah kelas khusus
P(H X)	= Probabilitas hipotesis H didasarkan pada kondisi X
P(H)	= Probabilitas hipotesis H
P(X H)	= Probabilitas hipotesis X didasarkan pada kondisi H
P(X)	= Probabilitas X

Metode TF-IDF merupakan metode untuk memberikan bobot pada hubungan sebuah kata (term) dengan sebuah dokumen. Metode ini menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan suatu kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut. Bobot hubungan antara sebuah kata dan dokumen akan tinggi jika frekuensi kata dalam dokumen juga tinggi dan frekuensi keseluruhan dokumen yang mengandung kata rendah dalam dokumen. Secara rumus, bobot TF-IDF dapat dihitung menggunakan persamaan sebagai berikut:

$$TF.IDFstd(t) = t\frac{f}{d} \times \log \frac{N}{dft} \quad (2)$$

Dimana t menunjukkan berapa kali istilah t muncul dalam dokumen d . N menunjukkan jumlah total dokumen dalam korpus. dft menunjukkan jumlah dokumen di mana istilah t terjadi [19]. Algoritma NBC dapat melakukan proses pengolahan data diskrit dan data kuantitatif dengan menggunakan sampel yang relative sedikit dan juga perhitungan pada algoritma NBC lebih cepat [20].

3.2. Bahan Penelitian

Dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksinasi covid-19 ini, materi utama berupa data tweet dan retweet di media sosial Twitter yang berkaitan dengan topik program vaksinasi. penelitian analisis sentimen di media sosial dengan studi kasus vaksinasi covid-19 ini menggunakan data tweet Twitter dengan kata kunci "Antivaksin Covid" atau dengan mengambil data berdasarkan akun yang terkait dengan berita tentang program vaksinasi seperti @KemenkesRI. Data tersebut dikumpulkan dari periode bulan Agustus 2021 hingga Desember 2021.

3.3. Alat Penelitian

Alat yang digunakan dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 berupa laptop dengan spesifikasi yang cukup untuk menjalankan proses pengolahan data dan mampu terhubung ke Internet. Sistem operasi dan program aplikasi yang digunakan dalam pengembangan aplikasi ini adalah sebagai berikut:

1. OS: Windows 10 64-bit
2. Bahasa Pemrograman: Python 3.10.2
3. Microsoft Excel 2019
4. Jupyter Notebook

3.4. Jalan Penelitian

Dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 ini, tahap pertama adalah penggunaan bahasa pemrograman Python untuk mengambil dan memvisualisasikan data yang ditampilkan di Microsoft Office Excel menggunakan library Python. Gambar 1 adalah jalan penelitian :

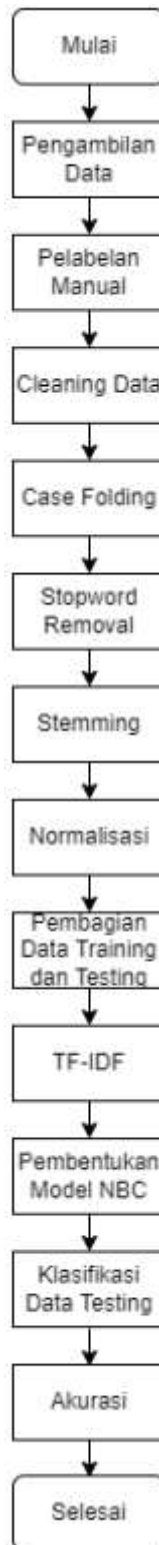


Fig. 1. Jalan Penelitian

3.4.1. Pengambilan Data

Pada penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 ini bahan utama yaitu berupa data tweet dan retweet di media sosial Twitter yang berkaitan tentang tema Program Vaksinasi. Penelitian ini menggunakan data tweet yang didapatkan dari Twitter dengan keyword/hastag “Antivaksin Covid” atau dengan pengambilan data berdasarkan akun-akun terkait berita mengenai program Vaksinasi seperti @KemenkesRI. Pengumpulan data dilakukan pada periode Bulan Agustus 2021- Desember 2021. Proses ini dilakukan dengan menggunakan software Anaconda dengan bantuan tools Jupyter Notebook. Proses selanjutnya menampilkan hasil data yang sudah berhasil di ambil, data tersebut akan ditampilkan ke dalam format csv atau Microsoft Office Excel.

Data tweet di ambil dari periode 1 Agustus 2021- Desember 2021, dengan jumlah data total 889. Data yang sudah berhasil diambil kemudian disimpan kedalam format Microsost Office Excel atau dalam bentuk format CSV yang ditunjukkan pada tabel 1 merupakan contoh data tweet yang diambil.

Table 1. Contoh data tweet dan re-tweet

No	Tweet
1	Pria berusia 33 tahun di Belgia, yang menjadi joki vaksinasi COVID-19, ditangkap saat hendak menerima vaksin kesembilan untuk wakili orangorang antivaksin. https://t.co/rfBIDOOsbo
2	Ponakan gue usia 9th bolak balik di suruh vaksin oleh pihak sekolah, bkn antivaksin sih ponakan2 & anak2 gue semua vaksin komplit baik yg wajib dr pemerintah maupun tambahan tp utk covid ini msh blm sreg
3	gerakan Antivaksin Covid sebagai salah satu ancaman kesehatan global. Terkait munculnya gerakan ini di Indonesia, Dicky menyarankan pemerintah meneliti lebih jauh penyebabnya di masing-masing daerah sehingga solusinya bisa disesuaikan.
4	Pemerintah terus menggenjot capaian vaksinasi COVID-19 di dalam negeri seiring dengan munculnya varian baru virus corona, omicron. Namun, akselerasi program vaksinasi terhambat oleh sejumlah hal, termasuk adanya gerakan antivaksin. https://t.co/o2kriVInG7
5	Pernah Dilabel Kelab Antivaksin, Man Utd Kini Diancam Penularan Wabak Covid-19

3.4.2. Pelabelan Data

Pada tahap pelabelan manual, tahap ini merupakan proses memberikan level terhadap kata pada dokumen sehingga dapat dianalisis lebih lanjut mengenai sifat yang positif atau negatif. Pada proses pelabelan manual peneliti membagi data tweet yang berisi tentang anti vaksin dengan data training label positif dan label negatif. Data tweet yang sudah dilabeli dengan jumlah data 889 tweet dengan masing-masing 277 tweet positif dan 612 tweet negatif dari data training. Pelabelan manual ditunjukkan pada gambar 2 bahwa label positif diberi kelas 1 dan nilai kelas 0 untuk label negatif.

	tanggal	akun	tweet	cleaned_tweet	lowered_tweet	removed_stopwords	stemmed_tweet	normalized_tweet	Label
0	2021-12-30 15:55:13	SyazaAufa	@fedtriyahya effect2 yg duk tulis ni nanti gen...	effect yg duk tulis ni nanti geng antivaksin ...	effect yg duk tulis ni nanti geng antivaksin ...	effect yg duk tulis ni geng antivaksin amik le...	effect yg duk tulis ni geng antivaksin amik le...	effect yang duk tulis ni geng antivaksin amik ...	0
1	2021-12-24 11:28:49	antaranews	Pria berusia 33 tahun di Belgia, yang menjadi ...	Pria berusia tahun di Belgia yang menjadi joki...	pria berusia tahun di belgia yang menjadi joki...	pria berusia tahun belgia menjadi joki vaksina...	pria usia tahun belgia jadi joki vaksinasi cov...	pria usia tahun belgia jadi joki vaksinasi cov...	0
2	2021-12-23 13:30:11	Rizm85	Ponakan gue usia 9th bolak balik di suruh fuck...	Ponakan gue usia th bolak balik di suruh fucks...	ponakan gue usia th bolak balik di suruh fucks...	ponakan gue usia th bolak balik suruh fucksin ...	ponakan gue usia th bolak balik suruh fucksin ...	ponakan saya usia th bolak balik suruh fucksin...	0
3	2021-12-21 01:42:57	voaindonesia	gerakan antivaksin COVID-19 sebagai salah satu...	gerakan antivaksin COVID sebagai salah satu an...	gerakan antivaksin covid sebagai salah satu an...	gerakan antivaksin covid salah satu ancaman ke...	gera antivaksin covid salah satu ancam sehat g...	gera antivaksin covid salah satu ancam sehat g...	0
4	2021-12-20 19:17:49	FirmRei	Gerakan Antivaksin COVID-19 Bisa Menjadi Ancam...	Gerakan Antivaksin COVID Bisa Menjadi Ancaman ...	gerakan antivaksin covid bisa menjadi ancaman ...	gerakan antivaksin covid menjadi ancaman keseh...	gera antivaksin covid jadi ancam sehat global	gera antivaksin covid jadi ancam sehat global	0

Fig. 2. Hasil Pelabelan manual

3.4.3. Cleaning data

Pada proses cleaning data, proses ini berfungsi untuk menghilangkan link url, tanda baca, angka, simbol dan username yang berada pada tweet dan re-tweet. Hasil akhir dari data yang telah di bersihkan seperti pada tabel 2.

Table 2. Data tweet bersih

No	<i>Tweet</i>
1	Pria berusia tahun di Belgia yang menjadi joki vaksinasi COVID ditangkap saat hendak menerima vaksin kesembilan untuk wakili orangorang antivaksin
2	Ponakan gue usia th bolak balik di suruh vaksin oleh pihak sekolah bkn antivaksin sih ponakan amp anak gue semua vaksin komplit baik yg wajib dr pemerintah maupun tambahan tp utk covid ini msh blm sreg
3	gerakan antivaksin COVID sebagai salah satu ancaman kesehatan global Terkait munculnya gerakan ini di Indonesia Dicky menyarankan pemerintah meneliti lebih jauh penyebabnya di masingmasing daerah sehingga solusinya bisa disesuaikan
4	Pemerintah terus menggenjot capaian vaksinasi COVID di dalam negeri seiring dengan munculnya varian baru virus corona omicron Namun akselerasi program vaksinasi terhambat oleh sejumlah hal termasuk adanya gerakan antivaksin
5	Goblokemang kamu punya empati sbg sekte antivaksinKeyakinan semu doang ditunjang rasa eksklusifbangga akan kelebihan dirimu amp org lain rendah Elon Musk kaya raya klaim dia tdk akan terpapar Covidmaklum OKmenolak lockdown tpi minta hrs ada karantina

3.4.4. Case Folding

Proses pengambilan data tweet “Antivaksin Covid” masih memiliki bentuk yang tidak standar. Jika data tersebut dianalisis maka hasil tidak akan akurat, sehingga diperlukan konversi dari bentuk awal menjadi bentuk standar.

Table 3. Hasil dari Case Folding

No	<i>Tweet</i>
1	pria berusia tahun di belgia yang menjadi joki vaksinasi covid ditangkap saat hendak menerima vaksin kesembilan untuk wakili orangorang antivaksin
2	ponakan gue usia th bolak balik di suruh vaksin oleh pihak sekolah bkn antivaksin sih ponakan amp anak gue semua vaksin komplit baik yg wajib dr pemerintah maupun tambahan tp utk covid ini msh blm sreg
3	gerakan antivaksin covid sebagai salah satu ancaman kesehatan global terkait munculnya gerakan ini di indonesia dicky menyarankan pemerintah meneliti lebih jauh penyebabnya di masingmasing daerah sehingga solusinya bisa disesuaikan
4	pemerintah terus menggenjot capaian vaksinasi covid di dalam negeri seiring dengan munculnya varian baru virus corona omicron namun akselerasi program vaksinasi terhambat oleh sejumlah hal termasuk adanya gerakan antivaksin
5	goblokemang kamu punya empati sbg sekte antivaksinkeyakinan semu doang ditunjang rasa eksklusifbangga akan kelebihan dirimu amp org lain rendah elon musk kaya raya klaim dia tdk akan terpapar covidmaklum okmenolak lockdown tpi minta hrs ada karantina

3.4.5. Stopword removal

Stopword removal yaitu melakukan penghapusan kata-kata yang memiliki informasi rendah dari sebuah teks (“yang”, “dan”, “di”, “dari” dll). Peneliti menggunakan proses Stopword Removal untuk menghapus kata-kata yang memiliki informasi rendah dari teks tweet “Anti Vaksin”. Tabel 5 menunjukkan hasil dari stopwords removal.

Table 4. Hasil dari Stopword removal

No	<i>Tweet</i>
1	pria berusia tahun belgia menjadi joki vaksinasi covid ditangkap hendak menerima vaksin kesembilan wakili orangorang antivaksin
2	ponakan gue usia th bolak balik suruh vaksin pihak sekolah bkn antivaksin sih ponakan amp anak gue semua vaksin komplit baik yg wajib dr pemerintah maupun tambahan tp utk covid msh blm sreg
3	gerakan antivaksin covid salah satu ancaman kesehatan global terkait munculnya gerakan indonesia dicky menyarankan pemerintah meneliti lebih jauh penyebabnya masingmasing daerah solusinya disesuaikan
4	pemerintah terus menggenjot capaian vaksinasi covid negeri seiring munculnya varian baru virus corona omicron akselerasi program vaksinasi terhambat sejumlah termasuk adanya gerakan antivaksin
5	goblokemang kamu punya empati sbg sekte antivaksinkeyakinan semu doang ditunjang rasa eksklusifbangga kelebihan dirimu amp org rendah elon musk kaya raya klaim tdk terpapar covidmaklum okmenolak lockdown tpi minta hrs karantina

3.4.6. Stemming

Stemming merupakan proses penghilangan infleksi kata menjadi bentuk dasar. Peneliti menggunakan Python Sastrawi dalam proses Stemming, hal ini dikarenakan data tweet yang berbahasa Indonesia semua kata tambahan sufiks dan prefix harus dihilangkan. Hasil stemming ditunjukkan pada tabel 5.

Table 5. Hasil dari Stemming

No	Tweet
1	pria usia tahun belgia jadi joki vaksinasi covid tangkap hendak terima vaksin sembilan wakil orangorang antivaksin
2	ponakan gue usia th bolak balik suruh vaksin pihak sekolah bkn antivaksin sih ponakan amp anak gue semua vaksin komplit baik yg wajib dr perintah maupun tambah tp utk covid msh blm sreg
3	gera antivaksin covid salah satuancam sehat global kait muncul gera indonesia dicky saran perintah teliti lebih jauh sebab masingmasing daerah solusi sesuai
4	perintah terus genjot capai vaksinasi covid negeri iring muncul varian baru virus corona omicron akselerasi program vaksinasi hambat jumlah masuk ada gera antivaksin
5	goblokemang kamu punya empati sbg sekte antivaksinkeyakinan semu doang tunjang rasa eksklusifbangga lebih diri amp org rendah elon musk kaya raya klaim tdk papar covidmaklum okmenolak lockdown tpi minta hrs karantina

3.4.7. Normalisasi

Normalization merupakan penyeragaman pada term yang mengalami kesalahan penulisan atau menggunakan bahasa yang tidak baku. Hasil dari proses Normalization berupa data tweet yang lebih terstruktur serta dapat dilakukan perhitungan pada proses selanjutnya dengan hasil ditunjukkan pada tabel 6.

Table 6. Hasil dari Normalisasi

No	Tweet
1	pria usia tahun belgia jadi joki vaksinasi covid tangkap hendak terima vaksin sembilan wakil orangorang antivaksin
2	ponakan saya usia th bolak balik suruh vaksin pihak sekolah bkn antivaksin sih ponakan amp anak saya semua vaksin komplit baik yang wajib dari perintah maupun tambah tapi untuk covid masih blm sreg
3	gera antivaksin covid salah satuancam sehat global kait muncul gera indonesia dicky saran perintah teliti lebih jauh sebab masingmasing daerah solusi sesuai
4	perintah terus genjot capai vaksinasi covid negeri iring muncul varian baru virus corona omicron akselerasi program vaksinasi hambat jumlah masuk ada gera antivaksin
5	goblokemang kamu punya empati sbg sekte antivaksinkeyakinan semu doang tunjang rasa eksklusifbangga lebih diri amp orang rendah elon musk kaya raya klaim tdk papar covidmaklum okmenolak lockdown tpi minta hrs karantina

3.4.8. Pembagian Data Training dan Data Testing

Pada proses ini digunakan untuk membagi data training dan data testing menggunakan Data frame agar data dilakukan training dan testing pada data tweet dengan hasil pembagian data training dan testing ditunjukkan pada gambar 2.

	normalized_tweet	label
0	vaksin membunuhmeninggal efek sakit bawa iya c...	0
1	pakar kesihatan tuju kumpul antivaksin tanggun...	0
2	geng antivaksin ni memang sign and symptom tak...	0
3	benda senang je beza vaksin amp antivaksin pos...	0
4	perkara paling tak masuk akal sekarang nie ial...	0

Fig. 3. Hasil pembagian data training dan testing

3.4.9. TF-IDF

TF-IDF digunakan untuk memberikan bobot pada hubungan sebuah kata (term) dengan sebuah dokumen. Metode ini menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi

kemunculan suatu kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut. Kode perhitungan ditunjukkan pada gambar 4.

```
# TF-IDF
vectorizer = TfidfVectorizer()

# tf_idf data_train
X_train_TFIDF = vectorizer.fit_transform(X_train)

# tf_idf data test
X_test_TFIDF = vectorizer.transform(X_test)
```

Fig. 4. Kode perhitungan TF-IDF

3.4.10. NBC

NBC digunakan untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Pada proses perhitungan NBC ini menggunakan library `NB_classifier_Multinomial` untuk menjalankan proses perhitungan secara otomatis.

3.4.11. Testing

Pada tahap testing merupakan tahapan untuk mengetahui tingkat keakuratan pemodelan yang telah dibangun serta untuk memprediksi label atau kelas dari data uji yang tersedia. Sebelum melakukan testing pertama tama membuat file pickle untuk eksekusi data testing dari data yang digunakan adalah 223 tweet yang sudah dilakukan pelabelan secara manual dari data training yang berjumlah 666 tweet dan 223 tweet yang digunakan data testing mengambil dari total data 889 tweet.

3.4.12. Akurasi

Menghitung nilai akurasi digunakan untuk menguji kebenaran penelitian yang diteliti ditunjukkan pada gambar 5 menggunakan Naive Bayes.

```
result_tweet=[]
for i in range(len(predicted)):
    if(predicted[i]==1):
        sentiment_result='Positif'
    elif(predicted[i]==0):
        sentiment_result='Negatif'
    # result_tweet.append({'class':prediction_linear[i], 'result_nbc':sentiment_result})
    result_tweet.append({'Cleaned_Text':data_tweet[i], 'class':predicted[i] })
```

Fig. 5. Kode Pemanggilan Prediksi

Setelah mendapatkan nilai akurasi yang baik dari proses training dan testing dari pemodelan klasifikasi maka dari itu dilakukannya tahap klasifikasi untuk data keseluruhan agar mendapatkan hasil sentimen positif dan negatif yang telah diuji dari tahapan training dan testing. Pada gambar 6 merupakan hasil pengujian.

	normalized_tweet	label	predicted_label
0	adoii antivaksin nie kan duk kesah side effect...	0	0
1	pelik geng antivaksin ni kecil dulu kau tak cu...	1	0
2	facebook padam ratus akaun kempen antivaksin c...	1	0
3	aku ja antivaksin ni share pasai amal depa ceg...	0	0
4	mati covid tu kata salah hospital covid ni asa...	0	0

Fig. 6. Hasil Pengujian

Selanjutnya menghitung nilai akurasi dari data training ditunjukkan pada gambar 7.


```

Accuracy: 0.722%

              precision    recall  f1-score   support

 Positive     0.72         0.99         0.84         158
 Negative     0.80         0.06         0.11          65

 accuracy          0.72         0.72         0.72         223
 macro avg         0.76         0.53         0.47         223
 weighted avg      0.74         0.72         0.63         223

Confusion Matrix:
[[157  1]
 [ 61  4]]
    
```

Fig. 7. Hasil Akurasi data training

4. Hasil dan Pembahasan

Data tweet yang didapatkan dari Twitter menggunakan keyword/hastag “Antivaksin Covid atau dengan pengambilan data berdasarkan akun-akun terkait berita mengenai program Vaksinasi Covid-19 seperti @KemenkesRI. Pengumpulan data dilakukan pada periode Bulan Agustus 2021-Desember 202, dengan jumlah data 889 data.

Pada tahap training data yang digunakan sejumlah 666 tweet dengan masing-masing 212 positif dan 454 negatif yang sudah diberi label secara manual. Sedangkan untuk data testing jumlah data yang digunakan adalah 223 data yang sudah diberi label positif dan negatif. Hasil dari penelitian ini mendapatkan nilai akurasi sebesar 72% berdasarkan hasil perhitungan confusion matrix di tabel 7. Hasil klasifikasi total dari “Antivaksin Covid” mendapatkan nilai 289 ulasan Positif dan 600 Ulasan negatif.

Table 7. Confusion matrix data testing

Kelas Prediksi	Kelas Aktual	
	Positif	Negatif
Positif	157	1
Negatif	61	4

Maka dapat menghitung nilai akurasi pada data testing yang dimana sudah diketahui nilai TP = 157, TN = 61, FP = 1 dan FN = 4 dengan menggunakan rumis persamaan untuk accuracy. Berikut untuk mengetahui nilai accuracy pada data testing pada Tabel 8.

Table 8. Akurasi data Uji

Akurasi data training	Hasil
Akurasi	72%

Hasil prediksi pada data tweet keseluruhan yang sudah diketahui hasil positif dan negatif dapat dilihat pada Gambar 8.

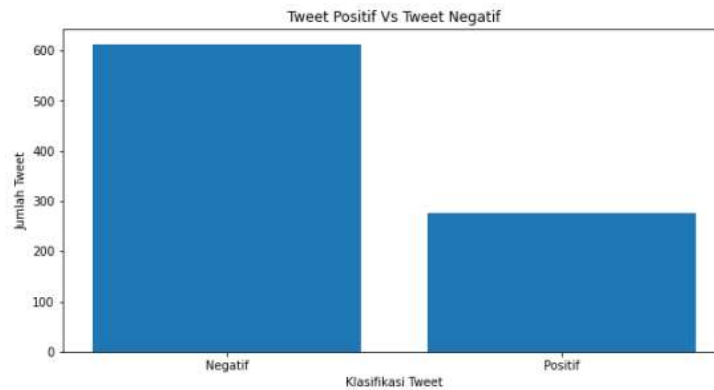


Fig. 8. Histogram data positif dan negatif

Pada sentimen negatif pengguna Twitter banyak yang berkomentar tentang tidak percaya pada vaksin, simpang siur informasi mengenai fakta-fakta yang terjadi lapangan. Takutnya masyarakat atas kejadian masyarakat yang meninggal secara mendadak setelah vaksin. Berikut merupakan contoh data tweet negatif pada Tabel 9.

Table 9. Contoh Data tweet negatif

No	<i>Tweet</i>
1	Pria berusia tahun di Belgia yang menjadi joki vaksinasi COVID ditangkap saat hendak menerima vaksin kesembilan untuk wakili orang-orang antivaksin
2	Gerakan Antivaksin COVID Bisa Menjadi Ancaman Kesehatan Global
3	Pemerintah Inggris Kecam Teror Kaum Antivaksin Covid di Gerbang Sekolah
4	Kumpulan anti vaksin lebih mempercayai kerajaan mempunyai muslihat tersembunyi sedangkan pihak kerajaan bertungkus lumus melindungi rakyat daripada ancaman COVID GetaranMY Kolum AntiVaksin
5	Goblokemang kamu punya empati sbg sekte antivaksinKeyakinan semu doang ditunjang rasa eksklusifbangga akan kelebihan dirimu amp org lain rendah Elon Musk kaya raya klaim dia tdk akan terpapar Covidmaklum OKmenolak lockdown tpi minta hrs ada karantina

Sedangkan sentimen Positif pada data tweet banyak yang membahas tentang perubahan dari masyarakat yang dulunya menolak vaksin berubah menjadi masyarakat yang provaksin. Berikut Contoh beberapa data tweet positif terdapat pada Tabel 10.

Table 10. Contoh data tweet positif

No	<i>Tweet</i>
1	cerita antivaksin amp tinggal covid jangan takut vaksin guys
2	Pemerintah terus menggenjot capaian vaksinasi COVID di dalam negeri seiring dengan munculnya varian baru virus corona omicron Namun akselerasi program vaksinasi terhambat oleh sejumlah hal termasuk adanya gerakan antivaksin
3	Tidak ada vaksin membunuhmeninggal karena efek penyakit bawaan iya seperti Covid rentan kematian tinggi kpd penderita yg punya penyakit bawaanseperti jantungdiabetes dll peduli amat dgn kelompok antivaksin sok otak dokter Abal Abal
4	Sepakat saya ga antivaksin dari lahir sampai SD di vaksin mau umroh divaksin Cuma untuk vaksin covid NO Krn seperti dipaksa dan diancam kebebasan saya
5	Meskipun bilangan yang enggan menerima vaksin COVID hanya antara lima hingga peratus namun sikap mereka boleh membawa mudarat kepada kelompok lebih besar selain berisiko menjadi pembawa virus itu antivaksin wacananya vaksin covid

5. Kesimpulan

Kesimpulan yang dapat ditarik setelah melakukan pengolahan dan pengujian data tweet pada media sosial Twitter adalah,

1. Akurasi yang dihasilkan adalah 72 %

2. Pemetaan sentimen positif dan sentimen negatif mengenai kelompok Anti-Vaksin dengan jumlah sebesar 289 berlabel positif dan 600 berlabel negatif. Sehingga sentimen akhir dalam klasifikasi kelompok Anti Vaksin pada penelitian ini bernilai “Sentimen Negatif”.

Beberapa saran yang dapat disampaikan mengenai hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Diharapkan pada proses preprocessing bisa mendapatkan data yang sesuai yang diinginkan seperti terdapat kata yang masih belum terbersihkan dengan benar.
2. Diharapkan pada penelitian selanjutnya terdapat sebuah pengembangan aplikasi yang dimana dapat mencari sebuah kata kunci yang ingin dilakukan analisis sentimen dengan jumlah data dan tanggal yang bisa di custom oleh user itu sendiri.

Daftar Pustaka

- [1] Adi Yahyadi. (2022). Analisis Sentimen Twitter Terhadap Kebijakan Ppkm Di Tengah Pandemi Covid-19. *Journal of Information System, Applied, Management, Accounting and Research.*, 6(2), 464–471. <https://doi.org/10.52362/jisamar.v6i2.791> [Accessed 25 June 2022]
- [2] Aldean, M. Y., Paradise, & Nugraha, N. A. S. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). 8106, 64–72.
- [3] Andika, L. A., Azizah, P. A. N., & Respatiwan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1), 34. <https://doi.org/10.13057/ijas.v2i1.29998> [Accessed 25 June 2022]
- [4] Aziz, A., Fauziah, F., & Fitri, I. (2021). Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 Menggunakan Metode Naive Bayes. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 842–851.. <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/381> [Accessed 25 June 2022]
- [5] Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, 1(1), 42–51. <https://doi.org/10.20895/dinda.v1i1.180> [Accessed 25 June 2022]
- [6] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *Smatika Jurnal*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455> [Accessed 27 June 2022]
- [7] Fitriana, F., Utami, E., & Al Fatta, H. (2021). Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 19–25. <https://doi.org/10.31603/komtika.v5i1.5185> [Accessed 27 June 2022]
- [8] Gafatia, I. W. D., & Hadinata, N. (2021). Analisis Pro Kontra Vaksin Covid 19 Menggunakan Sentimen Analysis Sumber Media Sosial Twitter. *Jurnal Pengembangan Sistem Informasi Dan Informatika*, 2(1), 34–42. <https://doi.org/10.47747/jpsii.v2i1.544> [Accessed 27 June 2022]
- [9] Junawan, H., & Laugu, N. (2020). Eksistensi Media Sosial, Youtube, Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia. *Baitul 'Ulum: Jurnal Ilmu Perpustakaan Dan Informasi*, 4(1), 41– 57. <https://doi.org/10.30631/baitululum.v4i1.46> [Accessed 27 June 2022]
- [10] Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(1), 32. <https://doi.org/10.24014/coreit.v7i1.12945> [Accessed 28 June 2022]
- [11] Lutfiyani, R. S., & Retnowati, N. (2021). Implementasi Pendeteksian Spam Email Menggunakan Metode Text Mining Dengan Algoritma Naive Bayes Dan Decision Tree J48. *Jurnal Komputer Dan Informatika*, 9(2), 244–252. <https://doi.org/10.35508/jicon.v9i2.5304> [Accessed 28 June 2022]

- [12] Mufidah, L., & Tejomurti, K. (2021). ANALISIS YURIDIS PELAKSANAAN PENGADAAN VAKSIN DALAM PENANGANAN CORONAVIRUS DISEASE 2019 (COVID-19) [A Legal Analysis of the Implementation of Vaccine Procurement in Managing the Coronavirus Disease 19 (Covid-19)]. *Law Review*, 3(March), 270. <https://doi.org/10.19166/lr.v0i0.3109> [Accessed 28 June 2022]
- [13] Ndruru, A. (2022). Analisis Sentimen UU Cipta Kerja Melalui Omnibus Law Menggunakan Naive Bayes Classifier (NBC) Dan Support Vector Machine (SVM). *Pelita Informatika: Informasi Dan ...*, 10, 85–90. <https://www.ejurnal.stmikbudidarma.ac.id/index.php/pelita/article/view/3768> [Accessed 28 June 2022]
- [14] Pinandhita, V. (2022). Update Corona RI 10 Agustus: Kasus Baru 5.926, Kasus Aktif Tembus 52 Ribu. <https://health.detik.com/berita-detikhealth/d6227143/update-corona-ri-10-agustus-kasus-baru-5926-kasus-aktif-tembus52-ribu> [Accessed 01 July 2022] Prasyaning Widayanti, L., & Kusumawati, E. (2021). HUBUNGAN PERSEPSI TENTANG EFEKTIFITAS VAKSIN DENGAN SIKAP KESEDIAAN MENGIKUTI VAKSINASI COVID-19. *HEARTY*, 9(2), 78. <https://doi.org/10.32832/hearty.v9i2.5400> [Accessed 01 July 2022]
- [15] Putri, R. N. (2020). Indonesia dalam Menghadapi Pandemi Covid-19. *Jurnal Ilmiah Universitas Batanghari Jambi*, 20(2), 705. <https://doi.org/10.33087/jiubj.v20i2.1010> [Accessed 01 July 2022]
- [16] Septianingrum, F., Jaman, J. H., & Enri, U. (2021). Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier. *Jurnal Media Informatika Budidarma*, 5(4), 1431. <https://doi.org/10.30865/mib.v5i4.3260> [Accessed 01 July 2022]
- [17] Wibowo, wahyu, & Novianti, wahyu ela. (2022). Analisis Sentimen Pengguna Twitter terhadap Program Kartu Prakerja di Tengah. 11(1).
- [18] Widowati, L. P., Manungkalit, E. M., Dewi, N. N. S. A., Suhaid, D. N., Wardani, D. W. K. K., & Ides, S. A. (2022). Percepatan Program Vaksinasi Covid-19 Demi Terwujudnya Herd Immunity di Wilayah Kerja Puskesmas Matraman, Jakarta Timur. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (Pkm)*, 5(4), 944–951. <https://doi.org/10.33024/jkpm.v5i4.5563> [Accessed 10 July 2022]
- [19] Yunus, N. R., & Rezki, A. (2020). Kebijakan Pemberlakuan Lock Down Sebagai Antisipasi Penyebaran Corona Virus Covid-19. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(3), 227–238. <https://doi.org/10.15408/sjsbs.v7i3.15083> [Accessed 10 July 2022]
- [20] Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422. <https://doi.org/10.30865/mib.v5i2.2845> [Accessed 10 July 2022]