

4_ANALISIS SENTIMENT PENGGUNA TWITTER MENGENAI

by Sariyun Naja Anwar

Submission date: 20-Apr-2023 04:10AM (UTC+0700)

Submission ID: 2069686262

File name: 4_ANALISIS_SENTIMENT_PENGGUNA_TWITTER_MENGENAI.pdf (1.38M)

Word count: 3262

Character count: 20078

ANALISIS SENTIMENT PENGGUNA TWITTER MENGENAI PROGRAM VAKSINASI COVID-19 MENGGUNAKAN ALGORITMA NAÏVE BAYES

Iwan Budianto, Sariyun Naja Anwar

Program Studi Teknik Informatika, Universitas STIKUBANK
Jl. Tri Lomba Juang Nomor 1, Kota Semarang
ibudianto4@gmail.com, sariyunna@edu.unisbank.ac.id

Abstract - The emergence of the covid-19 in 2020 in Indonesia had a major impact on social life. The government's effort to overcoming the spread of corona virus by implementing a covid-19 vaccination program so that the virus does not spread. Public opinions regarding the vaccination program are conveyed through social media, one of them is twitter. Twitter provides space for its users to express personal opinions regarding a topic, one of which is the vaccination program. To determine the tendency of public opinion towards the vaccination program, it is necessary to apply sentiment analysis. This research uses the naïve bayes algorithm to classify sentiment whether it's negative, positive, or neutral. Research result shows that the public tended to have a neutral opinion with a neutral percentage of 49,04%, negative 29,55%, and positive 21,41%. Sentiment classification shows that the naïve bayes algorithm is able to classify document with an accuracy value of 76%, precision 79%, recall 72%, and f-measure value 72%.

Keywords - Sentiment Analysis, Classification, Naïve Bayes, Twitter, Vaccination.

Abstrak - Munculnya virus covid-19 pada tahun 2020 di indonesia membawa dampak yang besar bagi kehidupan bermasyarakat. Upaya pemerintah dalam mengatasi penyebaran virus corona dengan menerapkan program vaksinasi covid-19 agar virus tidak semakin meluas. Berbagai opini masyarakat mengenai program vaksinasi disampaikan melalui media sosial, salah satunya twitter. Twitter memberi ruang kepada penggunanya untuk menyampaikan berbagai opini pribadi terkait suatu topik, salah satunya program vaksinasi. Untuk menentukan kecenderungan opini publik terhadap program vaksinasi, perlu dilakukan penerapan analisis sentimen. Penelitian ini menggunakan algoritma naïve bayes untuk mengklasifikasikan sentiment apakah termasuk negatif, positif, atau netral. Hasil penelitian menunjukkan masyarakat cenderung beropini netral dengan persentase netral 49,04%, negatif 29,55% dan positif 21,41%. Klasifikasi sentimen menunjukkan algoritma naïve bayes mampu mengklasifikasikan dokumen dengan nilai akurasi sebesar 76%, precision 79%, recall 72%, dan nilai f-measure 72%.

Kata Kunci - Analisis sentiment, Klasifikasi, Naïve Bayes, Twitter, Vaksinasi.

I. PENDAHULUAN

Pada bulan Desember tahun 2019, dunia digemparkan dengan kemunculan sebuah virus yang sangat berbahaya. Kemunculan virus ini pertama kali ditemukan di kota Wuhan, China dengan penyebaran yang sangat cepat. Virus yang diberi nama SARS-Cov-2 atau dikenal sebagai COVID-19 atau virus corona merupakan jenis virus yang menyerang sistem pernafasan manusia. Pada kasus tertentu, virus corona bisa mengakibatkan gangguan sistem pernafasan, infeksi pada paru-paru, hingga bisa menyebabkan kematian [1].

Pada tanggal 2 Maret 2020 merupakan awal ditemukannya virus SARS-Cov2 atau virus corona yang menjangkiti dua warga Jakarta [1]. Semenjak ditemukan kasus pertama kali itu, penyebaran virus corona di Indonesia mulai meluas. Semakin meluasnya penyebaran virus corona menyebabkan berbagai dampak buruk bagi masyarakat terutama dari segi kesehatan. Selain itu, dampak negatif juga dirasakan di bidang lain seperti sosial, pendidikan, dan ekonomi.

Pemerintah Indonesia berupaya untuk menekan penyebaran virus corona dengan menerapkan beberapa program, diantaranya menerapkan program pembatasan kegiatan masyarakat dan program vaksinasi. Kebijakan pemerintah didalam mengatasi pandemi COVID-19 dalam hal ini program vaksinasi memunculkan berbagai pendapat di masyarakat. Beragam pendapat yang disampaikan, banyak yang mendukung dan berkomentar baik, ada juga yang menolak dan berkomentar buruk terhadap kebijakan pemerintah.

Media sosial merupakan tempat bagi masyarakat untuk menyampaikan pendapatnya. Salah satu media sosial yang paling banyak digunakan di Indonesia yaitu Twitter. Pada tahun 2020, Pengguna aktif twitter di Indonesia mengalami lonjakan jumlah pengguna dari tahun sebelumnya yaitu 134 juta menjadi 166 juta pengguna aktif [2]. Banyaknya pengguna twitter memunculkan berbagai informasi yang dapat dimanfaatkan untuk mengetahui opini masyarakat terhadap program vaksinasi COVID-19 apakah cenderung bersifat negatif, positif, atau netral.

Opini masyarakat yang dipublikasikan melalui tweet terbukti menjadi sumber daya yang berharga bagi pemerintah. Dari data tweet yang mereka publikasikan, diperoleh data yang sangat banyak sehingga perlu dirubah dari data menjadi suatu informasi atau pengetahuan. Maka dari itu, diperlukan sebuah metode yang dapat membantu merubah data menjadi pengetahuan yaitu dengan menggunakan sentiment analisis.

Tanggapan sentiment atau yang biasa dikenal sebagai Sentimen analisis bertujuan untuk menganalisa, mengolah serta mengekstraksi data tertulis yang berupa pendapat dari sebuah entitas terhadap suatu topik sehingga memperoleh sebuah informasi atau pengetahuan [3]. Sentimen analisis menjadi sebuah metode yang banyak dipakai oleh para peneliti, yang bertujuan untuk menyajikan informasi dari sebuah himpunan data yang tidak terstruktur. Sehingga dari suatu penelitian, akan menghasilkan sebuah informasi yang memiliki makna seperti opini masyarakat, penilaian suatu kebijakan.

Algoritma *naïve bayes* adalah sebuah algoritma yang biasa digunakan sebagai pengklasifikasian teks. Pengklasifikasian *Naïve Bayes* menggunakan teknik machine learning yang memiliki performa yang cukup baik dalam jumlah data yang sangat besar. Namun algoritma ini memiliki kekurangan dalam hal sensitifitas pemilihan fitur, yang bisa mempengaruhi tingkat waktu perhitungan dan juga akurasi klasifikasi [4].

Penelitian bidang analisis sentiment telah banyak dilakukan para peneliti karena pembahasan yang menarik. Salah satunya penelitian yang menarik dilakukan oleh Juanita (2020) tentang pengklasifikasian persepsi masyarakat tentang pemilu tahun 2019. Data yang diambil dari twitter sebanyak 221 yang diklasifikasikan berdasarkan sentiment negatif, positif, dan netral. Analisis sentiment dilakukan dengan algoritma *Naïve Bayes* dengan tingkat akurasi yang didapat 76%.

Penelitian yang dilakukan Ratnawati (2018) yang membahas mengenai opini film yang diambil dari komentar di media sosial twitter. Hasil pengklasifikasian menggunakan algoritma *naïve bayes* memperoleh nilai akurasi 90%.

Penelitian yang akan dilakukan yaitu melakukan analisis sentiment netral, sentiment positif, serta sentiment negatif menggunakan algoritma *naïve bayes* terhadap program vaksinasi covid-19. Data diambil dari media sosial twitter.

A. Analisis Sentiment

Analisis sentiment ialah bagian dari *text mining* yang berfokus pada opini orang, sentiment, emosi dan tingkah laku terhadap suatu objek seperti layanan, produk, individu, organisasi serta atribut lainnya. Pengertian dari analisis sentiment yaitu proses mengolah dan mengekstraksi data yang berbentuk teks secara otomatis agar memperoleh informasi sentiment

yang terdapat dalam suatu kalimat tertentu. Tujuan dari analisis sentiment yaitu untuk mengetahui dan melihat kecenderungan opini atau pendapat terhadap suatu object oleh seseorang, apakah cenderung beropini negatif, positif atau netral[5]. Analisis sentiment dapat digunakan untuk memprediksi, menganalisa suasana publik atau masyarakat, menggambarkan kecenderungan suasana hati seseorang pada suatu topik tertentu karena analisis sentiment berfokus pada analisa dan pemahaman emosi dari tinjauan pola pada teks.[6].

B. Klasifikasi

Klasifikasi merupakan metode yang dapat digunakan dalam bidang data mining. Klasifikasi merupakan metode yang dapat menemukan suatu fungsi atau model dalam membedakan kelas tertentu yang bertujuan untuk memprediksi kelas label dari suatu objek labelnya yang belum diketahui. Klasifikasi melakukan pembuatan suatu model algoritma berdasarkan data latih yang tersedia, sehingga data yang baru atau data uji diklasifikasikan berdasarkan model tersebut. Diharapkan sebuah sistem mampu melakukan proses pengklasifikasian dataset dengan benar. Akan tetapi kinerja sistem tidak dapat melakukan semua klasifikasi dengan benar, sehingga perlu dilakukan pengukuran kinerja sistem dengan matriks konfusi [7].

C. Pembobotan Kata

Pembobotan kata merupakan tahapan menghitung dan menentukan bobot suatu kata sehingga dapat meningkatkan kinerja analisis sentimen suatu algoritma dalam pengklasifikasian teks[6]. Penelitian ini menggunakan teknik pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*). TF-IDF merupakan penggabungan dua metode untuk memberikan bobot suatu kata, yaitu dengan menghitung nilai TF kemudian mengalikannya dengan nilai IDF. Perhitungan TF-IDF menggunakan *library* sklearn dari program python. Perhitungan TF-IDF dapat diketahui pada persamaan (1),(2) dan (3)[8]

$$TF(d,t) = f(d,t) \quad (1)$$

$$IDF(t) = \log \left(\frac{1+Nd}{1+df(t)} \right) + 1 \quad (2)$$

$$TF - IDF = TF(d,t) \times IDF(t) \quad (3)$$

Dimana :

$TF(d,t)$: nilai term frequency dalam dokumen d

$f(d,t)$: Frekuensi term t dalam dokumen d

$IDF(t)$: Inverse document frequency dari term t

Nd : Jumlah keseluruhan dokumen

$df(t)$: Jumlah dokumen yang mengandung term t

D. Naïve Bayes

Algoritma *naïve bayes* merupakan algoritma yang sering dipakai dalam pengklasifikasian teks. Pengklasifikasian *naïve bayes* merupakan sebuah algoritma yang mengklasifikasikan teks dengan mempertimbangkan nilai probabilitas statik pada kelas tertentu sehingga pola tersebut digunakan untuk

memprediksi keanggotaan kelas pada suatu data uji berdasarkan prediksi probabilitas dari keanggotaan kelas data latih [9]. Penggunaan algoritma *naïve bayes* banyak ditemukan pada teknik klasifikasi teks di twitter dengan menggunakan metode seperti *maximum entropy classification*, *multinomial naïve bayes*, dan *unigram naïve bayes*. [10]. Algoritma *naïve bayes* dapat dirumuskan dengan persamaan (4) [9]

$$P(c|X) = \frac{P(X|c) \cdot P(c)}{P(X)} \quad (4)$$

Dimana :

- P(c|X) probabilitas hipotesis berdasarkan kondisi
- P(X|c) probabilitas berdasarkan kondisi pada hipotesis
- P(c) probabilitas hipotesis
- P(X) probabilitas c

Multinomial naïve bayes merupakan model dari algoritma *naïve bayes* yang biasa dipakai untuk pengklasifikasian teks. *Multinomial naïve bayes* tergolong dalam metode *supervised learning* sehingga data yang digunakan terlebih dahulu dilakukan pelabelan sebelum tahapan pelatihan data. Persamaan model *multinomial naïve bayes* untuk menghitung probabilitas dari dokumen *d* berada di kelas *c* (5)[8]

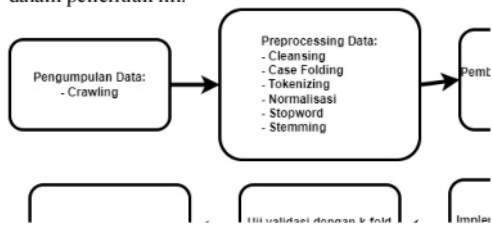
$$P(c|d) \propto P(c) \prod_{k=1}^n P(t_k|c) \quad (5)$$

dimana :

- P(c|d) Probabilitas dokumen d pada kelas c
- P(c) Prior probability suatu dokumen pada kelas c
- {t1,t1,t1,...,tn} Token dalam dokumen d yang menjadi bagian dari *vocabulary* dengan jumlah n
- P(tk|c) Probabilitas bersyarat dari term *tk* di dokumen pada kelas c

II. METODE PENELITIAN

Pada Gambar 1 akan dijelaskan beberapa tahapan dalam penelitian ini.



Gambar 1. Alur penelitian

1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan metode *crawling* pada sosial media twitter dengan fasilitas API (*Application Programming Interface*) twitter. Proses *crawling* menggunakan kata kunci “vaksin”

atau “vaksinasi” untuk mengumpulkan data *tweet*. Pengambilan data dilakukan pada tanggal 25 Februari 2022 sampai 26 Maret 2022 dan data disimpan dalam format file *.csv*. Data yang didapatkan sebanyak 836 data yang kemudian data dilabeli secara manual untuk menentukan sentiment positif, netral dan negatif.

2. Preprocessing Data

Preprocessing data merupakan tahapan yang dilakukan untuk merubah data teks menjadi data yang lebih terstruktur. Proses ini diperlukan guna membantu algoritma dalam melakukan pembelajaran terhadap model[6]. Tahapan *preprocessing* data meliputi *cleansing*, *case folding*, *tokenizing*, normalisasi, *stopword*, dan *stemming*. Penjelasan detail tentang *preprocessing* data dapat dilihat pada Tabel 1

Tabel 1. Penjelasan *preprocessing* data

Proses	Penjelasan
<i>cleansing</i>	Proses membersihkan dokumen dari berbagai komponen yang tidak diperlukan seperti URL, hashtag(#), '@', koma (,), titik(.), angka, emoticon, dan tanda baca lainnya
<i>case folding</i>	Proses merubah semua karakter alphabet menjadi seragam yaitu kecil semua (<i>lower case</i>)
<i>tokenizing</i>	Proses memisahkan setiap kata dari kalimat penyusunnya
normalisasi	Proses merubah kata tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI)
<i>stopword</i>	Proses menghilangkan kata-kata yang dianggap tidak memiliki arti atau makna serta pengaruh apapun dalam analisis sentimen
<i>stemming</i>	Proses mengubah kata menjadi kata dasar berdasarkan kaidah bahasa Indonesia yang benar

3. Pembobotan Kata dengan TF-IDF

Proses pembobotan kata dilakukan dengan cara menghitung nilai atau bobot setiap kata sehingga masing-masing kata diketahui bobotnya. Proses pembobotan tersebut dilakukan dengan teknik pembobotan TF-IDF. Implementasi penghitungan TF-IDF dilakukan pada program python dengan menggunakan *library* *sklearn* berdasarkan persamaan (1), (2), dan (3).

4. Implementasi Algoritma *Naïve Bayes*

Setelah dilakukan *preprocessing* data dan pembobotan kata, tahapan selanjutnya yaitu implementasi sentiment analisis menggunakan algoritma *naïve bayes*. Penghitungan algoritma *naïve bayes* berdasarkan persamaan (5).

5. Uji Validasi

Untuk memperkirakan kinerja sesungguhnya dari *machine learning*, diperlukan suatu uji validasi. Penelitian ini menggunakan metode uji validasi dengan *K-Fold Cross Validation*. Proses validasi dengan *K-fold cross validation* terbukti dapat menghilangkan *noise* pada kata sehingga akurasi pada algoritma dapat meningkat.[6]. Pada metode ini terdapat mekanisme keacakan pada himpunan data sehingga akurasi rata-rata yang dihasilkan menjadi tidak konstan. Hal ini dapat dimanfaatkan untuk memperoleh angka *fold* terbaik berdasarkan akurasi yang dihasilkan dari suatu model klasifikasi [11]

6. Hasil Klasifikasi

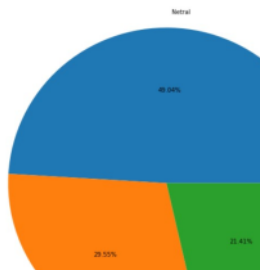
Setelah melalui tahapan *preprocessing* data, pembobotan kata, implementasi algoritma *naive bayes*, serta uji validasi dengan *K-fold cross validation* maka akan dihasilkan hasil klasifikasi yang akan dijelaskan dan ditampilkan pada bab selanjutnya.

III. HASIL DAN PEMBAHASAN

Proses sentiment analisis dalam penelitian ini memanfaatkan bahasa pemrograman python yang bisa diakses melalui situs www.colab.research.google.com. Proses tersebut dijalankan dengan tambahan beberapa *library* python yaitu *python-csv*, *pandas*, *numpy*, *sklearn*, dan *sastrawi*. Jumlah data yang dipakai pada sentiment ini sejumlah 836 data. Kemudian data dilabeli secara manual untuk menentukan kelas label tiap data. Data dibagi menjadi 3 kelas label, yaitu negatif, positif dan netral. Pembagian data berdasarkan kelas label ditunjukkan pada Tabel 2 dan Gambar 1.

Tabel 2. Pembagian data

Netral	Negatif	Positif	Jumlah
410	247	179	836



Gambar 2. Persentase pembagian data

Disajikan satu buah data *tweet* yang akan menjadi contoh dalam tahapan *preprocessing*. Data contoh diberi label T1. T1 = Aturan Baru Covid Singapura, Gua dan Warga Tolak Vaksin Harus Bayar RS Sendiri

<https://t.co/HzoE5aMQum>. Contoh penerapan *preprocessing* data ditunjukkan pada Tabel 3.

Tabel 3. Contoh penerapan *preprocessing* data

Tahapan	Hasil
<i>cleansing</i>	Aturan Baru Covid Singapura Gua dan Warga Tolak Vaksin Harus Bayar RS Sendiri
<i>case folding</i>	aturan baru covid singapura gua dan warga tolak vaksin harus bayar rs sendiri
<i>tokenizing</i>	'aturan', 'baru', 'covid', 'singapura', 'gua', 'dan', 'warga', 'tolak', 'vaksin', 'harus', 'bayar', 'rs', 'sendiri'
normalisasi	'aturan', 'baru', 'covid', 'singapura', 'aku', 'dan', 'warga', 'tolak', 'vaksin', 'harus', 'bayar', 'rs', 'sendiri'
<i>stopword</i>	'aturan', 'covid', 'singapura', 'warga', 'tolak', 'vaksin', 'bayar', 'rs'
<i>stemming</i>	'atur', 'covid', 'singapura', 'warga', 'tolak', 'vaksin', 'bayar', 'rs'

Tahapan selanjutnya yaitu melakukan pembobotan setiap kata dengan teknik TF-IDF. Sebagai contoh, disajikan tiga data yang mewakili masing-masing kelas sentiment, yaitu negatif, positif, dan netral serta data diberi label. Data dapat dilihat pada Tabel 4. Setiap kata akan dihitung bobotnya berdasarkan kemunculan frekuensi kata tersebut dalam dokumen. Penghitungan TF-IDF menggunakan persamaan (1), (2), dan (3). Hasil penghitungan TF-IDF ditampilkan pada Tabel 5.

Tabel 4. Contoh data untuk pembobotan kata

ID Dokumen	Label	Isi dokumen
d1	Negatif	badan sakit habis vaksin
d2	Netral	bentuk herd immunity babinsa damping warga vaksin
d3	Positif	alhamdulillah vaksin lengkap

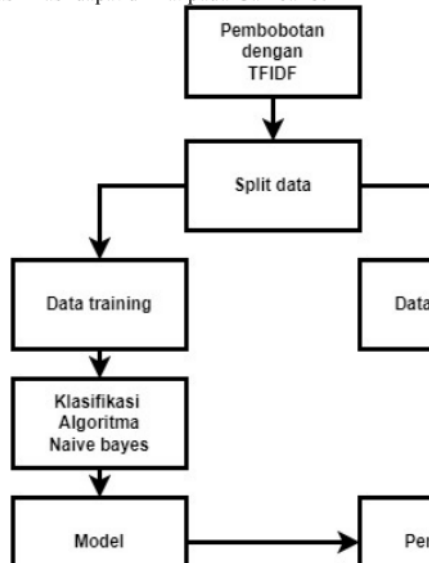
Tabel 5. Hasil penghitungan TF-IDF

Kata	TF-IDF		
	d1	d2	d3
alhamdulillah	0	0	0,599
babinsa	0	0,383	0
badan	0,514	0	0
bentuk	0	0,383	0
damping	0	0,383	0
habis	0,514	0	0
herd	0	0,383	0
immunity	0	0,383	0
lengkap	0	0	0,599
sakit	0,514	0	0
vaksin	0,454	0,339	0,530

warga	0	0,383	0
-------	---	-------	---

A. Klasifikasi Algoritma Naïve Bayes

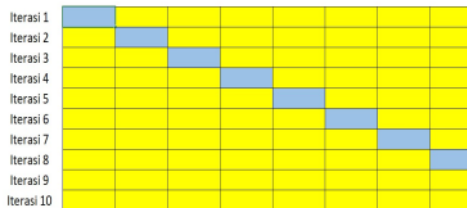
Data yang telah melewati tahapan pembobotan TF-IDF akan dibagi menjadi dua yaitu data *training* dan data *testing*. Data *training* berfungsi untuk membentuk model, sedangkan data *testing* berfungsi untuk menguji performa model sehingga diketahui nilai akurasi. Pembentukan model algoritma *naïve bayes* dari data *training* menggunakan persamaan (5). Pengujian model menggunakan uji validasi *K-fold cross validation* dengan nilai $k=10$. Tahapan klasifikasi dapat dilihat pada Gambar 3.



Gambar 3. Tahapan klasifikasi

B. Uji Validasi

K-fold cross validation merupakan metode proses uji validasi untuk memperkirakan kinerja sesungguhnya dari suatu model. Pengujian data menggunakan iterasi sebanyak 10 kali atau $k=10$. Data dipartisi menjadi 10 bagian, yang 9 bagian digunakan data *training* sedangkan 1 bagian digunakan untuk data *testing*. Data *testing* akan dilakukan perulangan dari iterasi 1 sampai iterasi 10. Proses iterasi pada metode *k-fold cross validation* dan pembagian data pada setiap iterasi ditampilkan pada Gambar 4 dan 5



Gambar 4. Iterasi 10-fold validation

Iterasi 1	84	84	84	84	84	84	83	83
Iterasi 2	84	84	84	84	84	84	83	83
Iterasi 3	84	84	84	84	84	84	83	83
Iterasi 4	84	84	84	84	84	84	83	83
Iterasi 5	84	84	84	84	84	84	83	83
Iterasi 6	84	84	84	84	84	84	83	83
Iterasi 7	84	84	84	84	84	84	83	83
Iterasi 8	84	84	84	84	84	84	83	83

Gambar 5. Pembagian data setiap iterasi

Hasil uji validasi dengan *K-fold cross validation* pada iterasi 1 sampai iterasi 10 ditampilkan pada Gambar 6 sampai Gambar 15

	precision	recall	f1-score
Negatif	0.80	0.67	0.73
Netral	0.76	1.00	0.87
Positif	1.00	0.50	0.67
accuracy			0.80
macro avg	0.85	0.72	0.75

Gambar 6. Iterasi-1

	precision	recall	f1-score
Negatif	0.79	0.55	0.65
Netral	0.65	0.97	0.78
Positif	1.00	0.42	0.59
accuracy			0.71
macro avg	0.81	0.65	0.67

Gambar 7. Iterasi-2

	precision	recall	f1-score
Negatif	0.68	0.62	0.65
Netral	0.63	0.91	0.75
Positif	0.81	0.45	0.58
accuracy			0.68
macro avg	0.71	0.66	0.66

Gambar 8. Iterasi-3

	precision	recall	f1-score
Negatif	0.81	0.71	0.76
Netral	0.78	0.90	0.84
Positif	1.00	0.78	0.88
accuracy			0.82
macro avg	0.86	0.80	0.83

Gambar 9. Iterasi-4

Hasil pengujian pada seluruh iterasi disajikan pada Tabel 6

	precision	recall	f1-score
Negatif	0.76	0.52	0.62
Netral	0.62	0.87	0.72
Positif	0.92	0.55	0.69
accuracy			0.69

Gambar 10. Iterasi-5

	precision	recall	f1-score
Negatif	0.77	0.63	0.69
Netral	0.69	0.92	0.79
Positif	1.00	0.65	0.79
accuracy			0.76

Gambar 11. Iterasi-6

	precision	recall	f1-score
Negatif	0.74	0.81	0.77
Netral	0.85	0.85	0.85
Positif	0.62	0.50	0.56
accuracy			0.80

Gambar 12. Iterasi-7

	precision	recall	f1-score
Negatif	0.81	0.79	0.80
Netral	0.87	0.90	0.88
Positif	0.75	0.60	0.67
accuracy			0.84

Gambar 13. Iterasi-8

	precision	recall	f1-score
Negatif	0.74	0.65	0.69
Netral	0.64	0.77	0.70
Positif	0.77	0.56	0.65
accuracy			0.69

Gambar 14. Iterasi-9

	precision	recall	f1-score
Negatif	0.83	0.77	0.80
Netral	0.71	0.83	0.76
Positif	0.85	0.65	0.73
accuracy			0.77

Gambar 15. Iterasi-10

Tabel 6. Hasil pengujian seluruh iterasi

K-Fold	Akurasi	Precision	Recall	f-measure
1	80%	85%	72%	75%
2	71%	81%	65%	67%
3	68%	71%	66%	66%
4	82%	86%	80%	82%
5	69%	77%	65%	68%
6	76%	82%	73%	76%
7	80%	74%	72%	72%
8	84%	81%	76%	78%
9	69%	72%	66%	68%
10	77%	79%	75%	77%
Rata-Rata	76%	79%	72%	72%

C. Hasil Klasifikasi

Hasil klasifikasi pada Tabel 6 memperoleh nilai rata-rata akurasi 76%, precision 79%, recall 72%, dan f-measure 72%. Nilai akurasi tertinggi terdapat pada iterasi ke-8 dengan nilai akurasi 84%. Sedangkan nilai akurasi terendah terdapat pada iterasi ke-3 dengan nilai akurasi 68%. Nilai precision tertinggi terdapat pada iterasi ke-4 dengan nilai precision 86%. Sedangkan nilai precision terendah terdapat pada iterasi ke-3 dengan nilai precision 71%. Nilai recall tertinggi terdapat pada iterasi ke-4 dengan nilai recall 80%. Sedangkan nilai recall terendah terdapat pada iterasi ke-3 dan 5 dengan nilai recall 65%. Nilai f-measure tertinggi terdapat pada iterasi ke-4 dengan nilai f-measure 82%. Sedangkan nilai f-measure terendah terdapat pada iterasi ke-3 dengan nilai f-measure 66%. Berdasarkan hasil klasifikasi tersebut, algoritma naïve bayes terbukti mampu melakukan analisis sentimen pada data teks dengan cukup baik. Hal ini ditunjukkan dengan perolehan hasil akurasi yang didapat yaitu sebesar 76%.

IV. KESIMPULAN

Hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa media sosial twitter merupakan media sosial yang dapat dimanfaatkan untuk memperoleh data opini publik berkaitan dengan suatu topik tertentu. Dari hasil crawling data, sentiment masyarakat tentang program vaksinasi covid-19 cenderung netral dengan nilai persentase 49,04%, sentiment negatif sebesar 29,55% dan sentiment positif sebesar 21,41%. Hasil klasifikasi membuktikan bahwa algoritma naïve bayes cukup bagus dalam analisis sentiment dengan data teks. Nilai akurasi yang dihasilkan sebesar 76%, precision 79%, recall 72%, dan f-measure 72%.

DAFTAR PUSTAKA

[1] Levani, Prasty, and Mawaddatunnadila, "Coronavirus Disease 2019 (COVID-19):

- Patogenesis, Manifestasi Klinis dan Pilihan Terapi,” *J. Kedokt. dan Kesehat.*, vol. 17, no. 1, pp. 44–57, 2021, [Online]. Available: <https://jurnal.umj.ac.id/index.php/JKK/article/view/6340>.
- [2] S. R. I. Rezeki, “Penggunaan Sosial Media Twitter dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi Dki Jakarta Dalam Penanganan Covid-19),” *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.
- [3] T. Yulianita, T. W. Utami, and M. Al Haris, “Analisis sentimen dalam penanganan covid-19 di indonesia menggunakan naive bayes classifier,” *Semin. Nas. Variansi*, pp. 235–243, 2020.
- [4] D. A. Muthia, “Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *Jurnalilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017.
- [5] I. Rozi, S. Pramono, and E. Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi,” *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [6] H. Tuhuteru and A. Iriani, “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier,” *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [7] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [8] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *J. Teknol. Informasi, Komputer, dan Apl. (JTika)*, vol. 2, no. 1, pp. 89–100, 2020, doi: 10.29303/jtika.v2i1.87.
- [9] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura,” *J. Buana Inform.*, vol. 11, no. 2, p. 140, 2020, doi: 10.24002/jbi.v11i2.3769.
- [10] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, “Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [11] T. T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation,” *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015, doi: 10.1016/J.PATCOG.2015.03.009.

4_ANALISIS SENTIMENT PENGGUNA TWITTER MENGENAI

ORIGINALITY REPORT

12%

SIMILARITY INDEX

12%

INTERNET SOURCES

3%

PUBLICATIONS

3%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

9%

★ www.jurnal.una.ac.id

Internet Source

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On