

IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM ANALISIS POLARISASI OPINI MASYARAKAT TERKAIT VAKSIN COVID-19

¹⁾ Abd. Charis Fauzan, ²⁾ Khoiril Hikmah

^{1,2)}Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar

^{1,2)}Jl. Masjid Nomor 22, Kota Blitar– Jawa Timur - Indonesia

E-mail : ¹abdcharis@unublitar.ac.id, ²avp.khoir@gmail.com

ABSTRAK

Hingga tahun 2022, Indonesia dilanda pandemi Covid-19. Covid-19 merupakan virus yang sangat mudah menyebar sehingga *World Health Organization* (WHO) menetapkan status *Virus Covid-19* sebagai pandemi global. Kasus *Covid-19* pertama kali terdeteksi di Indonesia pada tanggal 2 maret 2020. Sejak saat itu, banyak kasus yang terkonfirmasi positif. Pemerintah Indonesia sudah mengupayakan untuk menekan penyebaran *covid-19* agar dampak negatif yang ditimbulkan *covid-19* bisa dikendalikan, diantaranya dengan program vaksinasi *covid-19* yang terbagi menjadi beberapa tahapan, yaitu pemberian vaksin *covid-19* dosis satu, dosis dua, hingga vaksin booster. Akan tetapi, program vaksinasi *covid-19* untuk masyarakat memunculkan beragam opini di jagad media sosial, terutama twitter. Opini yang disampaikan cenderung terpolarisasi menjadi sentimen yang mendukung dan menolak. Untuk itu, penelitian ini bertujuan untuk mengetahui polarisasi opini masyarakat tentang program vaksinasi *Covid-19* menggunakan algoritma *naive bayes*. Proses dari analisis polarisasi opini meliputi pengumpulan data melalui twitter menggunakan *tools rapidMiner*, kemudian dilakukan *preprocessing* data dengan cara *case folding*, *tokenizing*, *filtering* dan *stemming*. Tahapan terakhir adalah melakukan klasifikasi opini masyarakat menggunakan algoritma *naive bayes*. Penelitian ini menghasilkan polarisasi opini masyarakat meliputi *tweets* sentimen positif sebesar 67%, *tweets* sentimen netral 9% dan *tweets* sentimen negatif 32%. Juga diperoleh tingkat akurasi dan *recall* masing-masing 88% dan 97%.

Kata Kunci: naive bayes, polarisasi opini, covid-19, vaksin

ABSTRACT

Until 2022, Indonesia was hit by the Covid-19 pandemic. Covid-19 is a virus that spreads very easily, so the World Health Organization (WHO) has declared the Covid-19 virus status as a global pandemic. The first case of Covid-19 was detected in Indonesia on March 2, 2020. Since then, many cases have been confirmed positive. The Indonesian government has made efforts to suppress the spread of Covid-19 so that the negative impacts caused by COVID-19 can be controlled, including the Covid-19 vaccination program which is divided into several stages, namely the administration of the Covid-19 vaccine, dose one, dose two, and the booster vaccine. However, the COVID-19 vaccination program for the community has raised various opinions in the social media universe, especially Twitter. Opinions expressed tend to be polarized into sentiments of support and rejection. For this reason, this study aims to determine the polarization of public opinion about the Covid-19 vaccination program using the Naive Bayes algorithm. The process of opinion polarization analysis includes collecting data via Twitter using the RapidMiner tools, then preprocessing the data by means of case folding, tokenizing, filtering and stemming. The last step is to classify public opinion using the Naive Bayes algorithm. This study resulted in polarization of public opinion including tweets of positive sentiment by 67%, tweets of neutral sentiment 9% and tweets of negative sentiment 32%. Also obtained the accuracy and recall rates of 88% and 97%.

Keyword: naive bayes, opinion polarization, covid-19, vaccine

PENDAHULUAN

Pada akhir tahun 2019, terdapat wabah *covid-19* di China tepatnya di Wuhan provinsi Hubei. Virus ini merupakan virus yang sangat mudah menyebar sehingga WHO (*World Health Organization*) menetapkan status *covid-19* sebagai pandemi global pada tanggal 11 februari 2020 [1].

Covid-19 masuk di Indonesia pertama kali

adanya warga yang terjangkit virus corona pada tanggal 2 maret 2020. Sejak saat itu, catatan dari satgas pemulihan *Covid-19*, banyak masyarakat yang terkonfirmasi positif *covid-19* sehingga angka *covid-19* di Indonesia semakin bertambah banyak dan menimbulkan dampak negatif diberbagai sektor, seperti sektor ekonomi dan pendidikan.[2].

Pandemi *covid-19* sudah berjalan selama 2 tahun hingga 2022, tetapi Indonesia masih ada

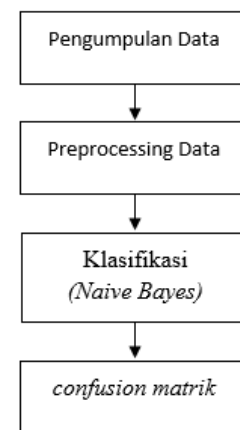
yang terpapar kasus *covid-19* pada update tanggal 13 Desember 2021 yang terkonfirmasi berjumlah 4.259.249 dengan jumlah positif 4.974, jumlah yang meninggal dunia 143.948, dan total sembuh mencapai angka 4.110.327 [3]. Pemerintah Indonesia sudah melakukan upaya untuk menekan angka penyebaran *Covid-19* agar dampak negatif yang ada dapat diminimalisir yaitu dengan upaya vaksinasi. Upaya vaksinasi *covid-19* mendapat berbagai macam opini dari masyarakat luas [2]. Opini yang diutarakan melalui media sosial twitter ada yang baik dan menerima, tetapi juga dan yang menolak. Opini masyarakat yang ada di media sosial *twitter* dapat dianalisis dengan berbagai metode analisis [2]. Metode analisis polarisasi opini adalah metode untuk menganalisis suatu data yang didapat dari internet. Dengan adanya analisis sentimen, polaritas opini dapat dikumpulkan, agar bisa memprediksi perasaan masyarakat bersifat positif atau negatif.

Topik analisis polarisasi opini ini banyak dilakukan oleh peneliti karena topik analisis ini sangat menarik untuk diteliti. Salah satunya penelitian yang dilakukan oleh [4] tentang analisis sentimen vaksin *sinovac* menggunakan Algoritma *Naive Bayes* mendapatkan *tweets* dengan sentimen positif sebanyak 86% dan *tweets* sentimen negatif sebanyak 14% dengan tingkat akurasi 92,96%. Metode *Naive Bayes* juga digunakan oleh [5] tentang isu vaksin *covid-19* menghasilkan akurasi 78% dan nilai *recall* 80%. *Naive Bayes* juga digunakan oleh peneliti sebelumnya dalam melakukan klasifikasi data berupa teks oleh [6] dengan topik pro kontra vaksin *covid-19* dengan hasil yang merespon positif sebanyak 8%. *Naive Bayes* adalah teks mining yang digunakan untuk klasifikasi sederhana yang memiliki tingkat akurasi tinggi. *Naive bayes* memiliki kelebihan yaitu jumlah data pelatihan yang kecil untuk memprediksi parameter yang

diperlukan dalam klasifikasi. *Naive Bayes* terbukti mempunyai tingkat kecepatan dan akurasi yang tinggi saat dipublikasikan kedalam basis data dengan jumlah yang besar. Analisis sentimen ini bertujuan untuk menentukan apakah vaksin *covid-19* bisa diterima atau tidak oleh masyarakat secara luas.

METODE

Jelaskan Pada penelitian ini, untuk mendapatkan hasil yang terbaik ada tahapan yang dilalui. Tahap yang dilalui terdiri dari pengumpulan data dan pelabelan data, selanjutnya tahap *preprocessing* data. Data yang sudah melakukan *preprocessing* data akan dikelompokkan dengan menggunakan algoritma *Naive Bayes*. Metode yang digunakan untuk proses pengujian yaitu confusion matrix Gambar 1 menunjukkan tahapan-tahapan yang dilakukan.



Gambar 1. Tahapan-tahapan penelitian

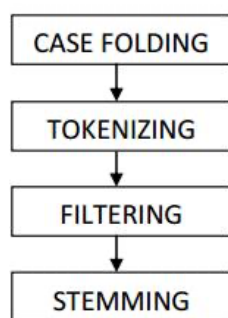
Pengumpulan data

Pengumpulan data dalam analisis sentimen berupa tanggapan masyarakat mengenai vaksin *covid-19*. Pada tahap pengumpulan data, peneliti melakukan dengan cara *crawling* data *twitter* menggunakan aplikasi rapidminer dengan kata kunci vaksin *covid-19*. Data yang digunakan berupa *tweets* berbahasa Indonesia. *Tweets* yang didapat berupa komentar-komentar masyarakat mengenai vaksinasi

covid-19. Data *twitter* yang telah berhasil didapat dari *twitter*, kemudian disimpan berbentuk file.csv.

Preprocessing data

Pada tahap *preprocessing* data akan diseragamkan bentuk dan format agar menjadi data yang dapat diolah pada tahap selanjutnya. *Preprocessing* terdiri dari *case folding*, *tokenizing*, *filtering* dan *stemming*. Gambar tahap *preprocessing* dapat dilihat pada Gambar 2.



Gambar 2 tahap *preprocessing* data

1. Case folding

Tahap ini merupakan proses penyeragaman huruf yang sebelumnya ada huruf yang besar dirubah menjadi huruf kecil semua. Contoh kalimat *Case folding input* “Ayo kita dukung dan sukseskan vaksinasi nasional tahap II untuk mencegah penyebaran covid-19” dan *output* “ayo kita dukung dan sukseskan vaksinasi nasional tahap ii untuk mencegah penyebaran covid-19”.

2. Tokenizing

Tahap ini merupakan tahap pemotongan kalimat menjadi bagian-bagian atau kata perkata. Contoh kalimat *tokenizing input* “Ayo kita dukung dan sukseskan vaksinasi nasional tahap II untuk mencegah penyebaran covid-19” dan *output* “Ayo”, “kita”, “dukung”, “dan”, “sukseskan”, “vaksinasi”, “nasional”, “tahap II”, “untuk”,

“mencegah”, “penyebaran”, “covid-19”.

3. Filtering

Tahap setelah tokenizing yaitu tahap menghilangkan kata yang muncul dengan jumlah banyak tetapi dianggap tidak memiliki arti.

4. Stemming

Tahap terakhir pada preprocessing yaitu pengubahan kata menjadi kata dasar yang benar. Contoh kalimat *Stemming input* “Ayo kita dukung dan sukseskan vaksinasi nasional tahap II untuk mencegah penyebaran covid-19” dan *output* “ dukung sukses vaksin nasional tahap II cegah sebar covid-19”.

Klasifikasi Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari data set yang diberikan. Algoritma menggunakan *teorema Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naive Bayes* memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya.

Naive Bayes pada penelitian ini digunakan untuk mengklasifikasi dokumen teks. Pada algoritma *Naive Bayes* setiap dokumen dipresentasikan dengan masukan atribut “ $a_1, a_2, a_3, \dots, a_n$ ” dimana a_1 adalah kata pertama dan berikutnya sampai a_n (kata ke- n), sedangkan V yaitu label kategori [9]. Selanjutnya yaitu mencari nilai tertinggi dari kategori teks yang diujikan (V_{MAP}) (McCallum, 1998). Hitung nilai probabilitas dengan menggunakan rumus :

$$P(W_k|V_j) = \frac{(n_k+1)}{(Jumlah\ Frekuensi + Jumlah\ Kata)} \quad (1)$$

Keterangan :

$P(W_k|V_j)$:

Probabilitas bobot kata sesuai kategori
 n_k : Nilai kemunculan frekuensi kata

sebagai contoh data artikel untuk kategori :

Positif : “Dukung perlambat pandemi dengan siap vaksin”

Netral : “Vaksinasi Covid-19 di Kudus tetap berlanjut saat ramadhan”

Negatif: “Tenaga kesehatan tolak vaksin Covid-19”

Maka jumlah kata yang didapat sebanyak 14 dan memiliki jumlah setiap frekuensi dari masing-masing kategori Positif = 5, jumlah frekuensi kategori netral = 6, jumlah frekuensi kategori negatif = 5.

Sebagai contoh dari kategori negatif hasil setelah preprocessing yaitu: “tenaga sehat tolak vaksin covid-19”. Dihitung nilai probabilitasnya:

$$P(\text{tenaga|negatif}) = (1+1) / (5+14) = 0.1052$$

$$P(\text{sehat|negatif}) = (1+1) / (5+14) = 0.1052$$

$$P(\text{tolak|negatif}) = (1+1) / (5+14) = 0.1052$$

$$P(\text{vaksin|negatif}) = (1+1) / (5+14) = 0.1052$$

$$P(\text{covid-19|negatif}) = (1+1) / (5+14) = 0.1052$$

Selanjutnya dimasukkan kedalam tabel probabilitas seperti Tabel 1.

Tabel 1. Probabilitas setiap kata

Kata	Probabilitas		
	Positif	Netral	Negatif
Tenaga	0.052	0.050	0.1052
Sehat	0.052	0.050	0.1052
Tolak	0.052	0.050	0.1052
Vaksin	0.105	0.10	0.105
Covid-19	0.052	0.10	0.105

Setelah mendapat nilai probabilitas kata pada setiap kategori, selanjutnya menghitung

probabilitas kategori dengan menggunakan rumus :

$$P(V_j) = \frac{\text{jumlah Dokumen setiap Kategori}}{\text{Total Dokumen}} \quad (2)$$

Diketahui

$$: \text{jumlah dokumen positif} = 1$$

$$: \text{jumlah dokumen netral} = 1$$

$$: \text{jumlah dokumen negatif} = 1$$

Jadi, probabilitas dari dokumen :

$$P(\text{positif}) = 1/3 = 0.33$$

$$P(\text{netral}) = 1/3 = 0.33$$

$$P(\text{negatif}) = 1/3 = 0.33$$

Confusion Matrix

Evaluasi yang akan digunakan pada analisis polarisasi opini disini menggunakan *confusion matrix*. *confusion matrix* adalah suatu tabel matriks, yang digunakan untuk mengukur suatu kinerja klasifikasi pada raangkaian data uji dengan nilai sebenarnya yang diketahui. *confusion matrix* memiliki informasi yang didalamnya ada perbandingan hasil dari sitem dengan hasil klasifikasi yang seharusnya.

Tabel 2 Rancangan Pemujian Model

	Kelas Prediksi	
	1	2
Kelas	1	2
Sebenarnya	1	2
	<i>TP</i>	<i>TN</i>
	<i>FP</i>	<i>FN</i>

Keterangan untuk tabel diatas dinyatakan sebagai berikut :

1. *True Positive* (TP), yaitu jumlah data dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
2. *True Negative* (TN), yaitu jumlah data dari kelas 0 yang benar dan diklasifikasikan sebagai kelas 0.

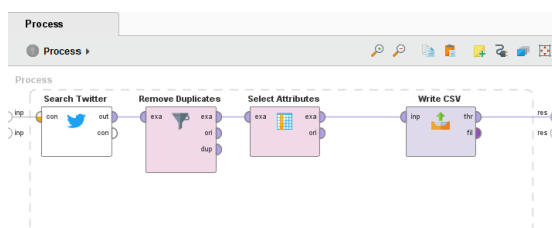
3. *True False* (TF), yaitu jumlah data deeari kelas 0 yang salah dan diklasifikasikan sebagai kelas 1.
4. *False Negative* (FN), yaitu jumlah data dari kelas 1 yang salah dan diklasifikasikan sebagai kelas 0

HASIL

Dalam penelitian analisis polarisasi opini ini, bahasa yang digunakan untuk menganalisis menggunakan bahasa pemrograman Php. Pelabelan dilakukan secara manual sesuai dengan kategorinya, yaitu positif, negatif atau netral. Selanjutnya proses *preprocessing* data dengan 4 tahap, yaitu *Case folding*, *Tokenizing*, *Filtering*, *Stemming*. Total data set yang akan digunakan sebanyak 210 *tweet* yang dibagi menjadi data latih sebanyak 180 dan data uji sebanyak 30. Data latih yang sudah melalui tahap *preprocessing* akan dijadikan bahan untuk proses klasifikasi data uji dengan algoritma Naive Bayes. Pengujian algoritma Naive Bayes akan menggunakan Confussion Matrix.

Hasil Pengumpulan Data

Pengumpulan data dilakukan dengan *crawling* data dengan aplikasi rapidminer dengan kata kunci “vaksin covid-19” dengan bahasa yang digunakan berupa bahasa Indonesia. Data yang didapat berupa komentar-komentar masyarakat mengenai vaksinasi covid-19. Jumlah data yang didapat sebanyak 721 data dengan data yang sesuai sebanyak 210 data. Pengumpulan data ini menggunakan akses *Twitter* API. Proses pengumpulan data bisa dilihat pada Gambar 3.



Gambar 3. Pengumpulan Data

Menggunakan RapidMiner

Hasil Preprocessing Data

Preprocessing data merupakan tahapan dimana data akan diseragamkan bentuk dan format supaya data dapat diolah pada tahap selanjutnya. *Preprocessing* terdiri dari *case folding*, *tokenizing*, *filtering* dan *stemming*.

Hasil proses Case Folding :

No	Kalimat
1	menkes budi gunadi sadikin yang menyatakan bahwa indonesia bisa menang lantaran telah amankan sejumlah dosis vaksin presiden jokowi dijadwalkan akan menerima vaksinasi covid dosis kedua besok pagi setelah kota gorontalo kabupaten gorontalo dan bone bolango giriran kabupaten gorontalo utara menerima pendistribusian vaksin harusnya yang divaksin bisa mencapai warga masyarakat di indonesia

Gambar 4 Hasil proses *Case Folding*

Gambar 4. menunjukkan hasil proses *case folding* dimana data yang sebelumnya masih ada kata berhuruf besar maka dalam proses *case folding* kata berhuruf besar dirubah menjadi huruf kecil semua.

Hasil Proses Tokenizing :

No	kata
1	menkes
2	budi
3	gunadi
4	sadikin
5	yang
6	menyatakan
7	bahwa
8	indonesia

Gambar 5 Hasil Proses *Tokenizing*

Pada proses *tokenizing* data yang sudah melalui proses *case folding* data dipotong-potong menjadi perkata. Seperti pada Gambar 5.

Hasil proses *Filtering* :

Show entries

No	kata
1	menkes
2	budi
3	gunadi
4	sadikin
5	indonesia
6	tenang
7	lantaran
8	amankan

Gambar 6. Hasil Proses *Filtering*

Pada tahap *filtering* data disaring, dimana kata-kata yang tidak diperlukan akan dibuang seperti kata konjungsi : yang, dan, bahwa, dengan dll.

Hasil proses *Stemming* Nazief & Adriani :

Show entries

No	kata
1	menkes
2	budi
3	gunadi
4	sadikin
5	indonesia
6	tenang
7	lantar
8	aman

Gambar 7 Hasil Proses *Stemming*

Proses *stemming* merupakan proses terakhir *preprocessing*, pada proses *stemming* data yang sudah melalui proses *filtering* data akan diubah menjadi kata dasar. Seperti kata lantaran menjadi kata lancar dan kata amankan menjadi kata aman.

Hasil Klasifikasi *Naive Bayes*

Setelah proses *preprocessing* dilalui dan mendapatkan data yang siap untuk diolah maka langkah selanjutnya yaitu implementasi *naive bayes* dilakukan. Proses implementasi *Naive Bayes* disini yaitu yang pertama data akan dihitung satu persatu katanya berapa persen kata itu masuk positif dan berapa persen kata itu

masuk negatif dengan cara kata ditambah 1 dibagi jumlah frekuensi ditambah jumlah seluruh kata maka akan didapat berapa persen kata tersebut. Dari hasil itu akan ditotal positif dan negatif kalau lebih banyak positif data akan masuk positif dan begitu juga dengan negatif.

The screenshot shows a web application titled "ANALISIS SENTIMEN" with a navigation bar including HOME, DATA LATH, DATA UJI, PREPROCESSING, NAIVE BAYES (highlighted), and PENGULJIAN. Below the navigation bar is a table with columns for No., Text \ Attr, and various sentiment categories (menkes, budi, gunadi, sadikin, indonesia, tenang, lancar, aman, dosis, vaksin). The table contains numerical values representing the classification results for a specific text input.

Gambar 8 Tahap Klasifikasi *Naive Bayes*

Evaluasi

Evaluasi adalah pengukuran performa terhadap klasifikasi yang telah dilakukan. Evaluasi disini menggunakan *confussion matrix*. Hasil *confussion matrix* dapat dilihat pada Gambar 9.

Hasil Evaluasi	
Evaluasi	Nilai
Akurasi	0.8848484848484848
Recall Positive	0.97297297297297
Precision Positive	0.87096774193548
Recall Negative	0.7037037037037037
Precision Negative	0.92682926829268

Gambar 9. Evaluasi *Confussion Matrix*

Berdasarkan hasil pengujian pada tabel 2, bahwa analisis algoritma *Naive Bayes* mendapatkan tingkat akurasi sebesar 88%, *recall* positif 97%, *precision* positif 87%, *recall* negatif 70% dan *precision* negatif 92%.

KESIMPULAN

Penelitian analisis polarisasi opini yang sudah dilakukan dapat menunjukkan sentimen masyarakat tentang vaksinasi *covid-19* yang diupaya pemerintah untuk menekan tingkat penyebaran *covid-19*. Hasil yang didapat dari analisis polarisasi opini menggunakan

algoritma *Naive Bayes* menghasilkan nilai sentimen positif 67%, negatif 32% dan netral 9%. Penelitian analisis sentimen ini dapat disimpulkan bahwa hasil sentimen positif lebih mendominasi dari pada sentimen negatif dan netral. Oleh karena itu, dapat dinilai adanya vaksinasi *covid-19* mendapat respons positif dari masyarakat

DAFTAR PUSTAKA

- [1] Wulandari, “Komunikasi Mui Provinsi Bengkulu Dalam Mensosialisasikan Vaksinasi Covid-19,” 2021, [Online]. Available: <http://repository.iainbengkulu.ac.id/6819/>.
- [2] B. Laurensz and Eko Sedyono, “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [3] T. K. K. P. C. V. Covid-19, “Angka Kesembuhan COVID-19 Meningkatkan Mencapai 4.110.327 Orang,” *13 Desember 2021*, 2021. <https://covid19.go.id/p/berita/angka-kesembuhan-covid-19-meningkat-mencapai-4110327-orang>.
- [4] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” 2021.
- [5] F. Septianingrum, J. H. Jaman, and U. Enri, “Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier,” vol. 5, pp. 1431–1437, 2021, doi: 10.30865/mib.v5i4.3260.
- [6] N. Hadinata, F. Teknik, I. Komputer, and U. B. Darma, “Analisis Pro Kontra Vaksin Covid 19 Menggunakan Sentiment Analysis Sumber Media Sosial Twitter,” vol. 2, no. 1, pp. 34–42, 2021.
- [7] S. H. Putra, *Merdeka Kreatif di Era Pandemi Covid-19*, Mesran. GREEN PRESS, 2020.
- [8] A. Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [9] I. A. Setiawan, T. H. P, and D. Nursantika, “Klasifikasi Artikel Berita Menggunakan Metode Text,” pp. 1–6, 2017.