

Analisis Sentimen Opini Publik Terhadap Program Vaksinasi Covid-19 Di Indonesia Pada Twitter Menggunakan Metode Naive Bayes Classifier

Priza Pandunata*, Kukuh Tri Winarno** Nugroho, Yanuar Nurdiansyah***, Nova El Maidah****

Fakultas Ilmu Komputer Universitas Jember

*priza@unej.ac.id, **kukuhtri99@gmail.com, ***yanuar_pssi@unej.ac.id, ****nova.pssi@unej.ac.id

ABSTRACT

The COVID-19 virus emerged in December 2019 in China and actively spread throughout the world including Indonesia in early 2020. Its spread is very fast and has caused millions of deaths. Therefore, the Indonesian government is actively holding a COVID-19 vaccination program to prevent the spread of the virus and make the public immune to the virus. But the program invites pros and cons among the community. Twitter is one of the social media that is famous for being a medium of opinion from the general public. The process of sentiment analysis can find and solve problems based on public opinion on social media such as Instagram. The classification method used in this research is Naive Bayes Classifier. The dataset can be obtained from data crawling process using Google Collabs and python programming language. The total dataset obtained is 2000. The data the labelled as positive, neutral, or negative. The labelling process result showed 1579 positive data, 277 negative data, and 144 neutral data. Then pre-processing is carried out on the data that has been labeled before, also word weighting process using TF-IDF. After that modelling is carried out using Naive Bayes Classifier and the last process is evaluation-testing. The high accuracy of the result from fourth experiment which compare 90% data training with 10% data testing produce 86% accuracy. While the result of sentiment test show that positive sentiment more than negative sentiment and neutral sentiment.

Keyword: *Sentiment analysis, COVID-19, Vaccine, TF-IDF algorithm, naive bayes classifier method.*

1. Pendahuluan

Virus COVID-19 adalah sebuah coronavirus yang ditemukan di China pada Desember 2019 dan aktif menyebar keseluruh dunia termasuk Indonesia pada sepanjang tahun 2020. Kemampuan penyebarannya yang sangat cepat dan mematikan dapat membuat kasus penyebaran sebanyak 187 juta dan diantaranya menyebabkan kematian sebanyak 2,5 juta jiwa pada Juli 2021 [1]. Oleh karena itu, Pemerintah Indonesia aktif mengadakan program vaksinasi untuk memutus rantai penyebaran COVID-19. Vaksin COVID-19 dikatakan telah melalui proses yang sangat lama dan diuji efektifitasnya secara ketat dan menyeluruh untuk memastikan masyarakat aman dan sehat setelah divaksinasi. Ada berbagai macam jenis dan tipe vaksin yang telah disediakan oleh pemerintah Indonesia. Setiap jenis vaksin terdiri dari bahan dasar dan tingkat efektifitas yang berbeda [2]. Program vaksin yang diadakan pemerintah Indonesia ini kontroversial di mata masyarakat. Beberapa orang mempercayai vaksin adalah solusi yang sangat tepat dan aman untuk memerangi virus COVID-19 dan dengan senang hati menerimanya. Sementara masyarakat lainnya masih meragukan manfaat dan kondisi tubuh setelah divaksinasi. Kebingungan ini diyakini disebabkan oleh berbagai informasi yang sudah menyebar luas dan menjadi obrolan panas diantara masyarakat. Organisasi Kesehatan Dunia (WHO) sebenarnya sudah menyatakan keyakinannya pada efektifitas vaksin COVID-19 [3].

Menggunakan media sosial sudah jadi kegiatan wajib sehari-hari untuk masyarakat seluruh, salah satunya Indonesia. Satu dari sekian banyak media sosial yang terkenal adalah Twitter (Perdana, A., et al, 2022). Twitter selalu menjadi sarana bagi banyak masyarakat Indonesia untuk mengungkapkan pikiran, pendapat, opini, keluhan, dan saran mereka masing-masing dalam sebuah fitur yang disebut *tweet* [4]. Banyak informasi yang bisa didapatkan dari seseorang dengan melihat isi dari *tweet* yang diposting oleh seseorang tersebut. Sebuah *tweet* bisa sangat mudah untuk tersebar luas dikalangan masyarakat dan bahkan terkadang bisa masuk dalam perhatian *public figure* atau acara TV sehingga membuat bahasan *tweet* tersebut viral atau terkenal dalam beberapa saat. Data yang ada didalam media sosial *Twitter* ini dapat digunakan untuk membantu dalam melaksanakan sebuah penelitian analisis sentimen [5].

Diperlukan analisis sentimen pada media sosial mengenai hal ini dengan tujuan untuk mengetahui penyebab pro kontra masyarakat dalam menanggapi program vaksin ini. Analisis sentimen atau Opinion Mining adalah metode mencari solusi dari sebuah masalah berdasarkan opini publik atau masyarakat yang disebarkan dalam bentuk teks di media sosial [6]. Analisis sentimen berguna bagi pemerintah dan organisasi

bisnis untuk mengungkapkan opini dan opini publik tentang kebijakan dan peraturan yang baru atau yang akan diterapkan tanpa harus menggunakan pilihan metode penelitian tradisional. Analisis sentimen mengklasifikasikan atau mengelompokkan polaritas teks dalam *tweet* untuk menemukan bahwa pendapat atau opini yang ada didalamnya bersifat negatif, netral, atau positif. Analisis sentimen juga biasa digunakan dalam membantu penelitian tentang metode klasifikasi seperti *Naive Bayes Classifier* [7]. Metode *Naive Bayes Classifier* merupakan algoritma yang umum digunakan dengan maksud menentukan kemungkinan nilai tertinggi untuk mengelompokkan data uji dan memasukkannya di sebuah kategori. Metode ini merupakan algoritma pengelompokan yang dibuat dengan teorema *bayes*. Algoritma ini menentukan gambaran umum opini publik dengan mengklasifikasikannya menggunakan teknik probabilistik dan statistik. Secara khusus, ini ditandai dengan memprediksi peluang berdasarkan pengalaman dan peristiwa masa lalu atau memiliki asumsi yang kuat [8].

Penelitian ini berfokus kepada pernyataan opini masyarakat mengenai program vaksinasi *COVID-19* di Indonesia pada Twitter yang mengandung sentimen positif, negatif, dan netral. Uji evaluasi juga dilaksanakan untuk menguji model klasifikasi yang telah dihasilkan. Penelitian ini diharapkan dapat membantu pemerintah dengan mengetahui sentimen dan opini masyarakat Indonesia terhadap program pemerintah mengenai pembagian vaksinasi *COVID-19* sehingga bisa menjadi sebuah pemikiran dan pertimbangan atau referensi untuk menyusun ulang strategi dalam memperbaiki pelaksanaan program ini baik dimasa kini maupun di masa mendatang apabila terjadi pandemi serangan virus yang serupa di Indonesia.

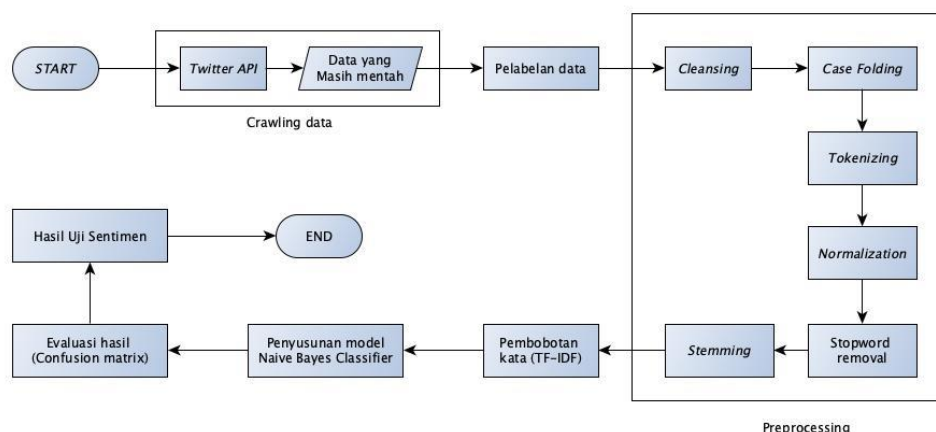
2. Metode Penelitian

2.1. Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini dilakukan dengan mengambil *tweet* dari Twitter menggunakan *tools* Google Collabs dan dikombinasikan dengan bahasa pemrograman *python* untuk *crawling data* dengan mencari *tweet* yang memiliki 10 kata kunci yang paling tren sepanjang tahun 2021 yaitu #COVID19, #vaksin, #VaksinasiCovid19, #vaksinasicovid19, #corona, ayo vaksin, jangan vaksin, vaksin covid, covid-19, vaksinasi covid. Waktu pengambilan data *tweet* dilaksanakan selama sekitar 2 bulan yaitu dimulai dari tanggal 8 Agustus 2021 sampai dengan 1 Oktober 2021. Total data *tweet* yang diambil adalah sebanyak 2000 data.

2.2. Tahapan Penelitian

Tahapan penelitian merupakan urutan langkah-langkah yang dilakukan dalam penelitian. Tahapan dalam penelitian ini dimulai dengan pengumpulan data mentah, pemrosesan data mentah menjadi *dataset*, penyusunan model, dan hasil akhir penelitian. Alur tahapan penelitian selengkapnya dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Dapat dilihat dari Gambar 1 bahwa tahapan penelitian diawali dengan proses *crawling data* dengan kode pemrograman *python* yang dijalankan dengan menggunakan Google Collabs dan menghasilkan kumpulan

data mentah. Selanjutnya kumpulan data mentah akan dilabeli oleh Ahli Bahasa dengan menggunakan teknik *Natural Processing Language* [9] berbasis konteks sebagai standar pelabelan data mentah kedalam kelompok sentimen positif atau negatif atau netral. Selanjutnya data mentah akan diolah agar menjadi data yang matang atau *dataset* dengan tujuan mengurangi *noise* dan meningkatkan kualitas data menggunakan *text preprocessing* yang terdiri dari beberapa tahapan didalamnya yaitu *cleansing*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Tahap selanjutnya adalah pembobotan kata dengan memanfaatkan algoritma *Term Frequency – Inverse Document Frequency (TF-IDF)*. *TF-IDF* terdiri dari *Term Frequency* (banyaknya suatu kata pada suatu data) dan *Inverse Document Frequency (Inverse dari Document Frequency yang berarti banyaknya suatu dokumen yang mengandung suatu kata)*. Perhitungan *TF-IDF* dapat dilihat pada Persamaan 1.

$$W_{t,d} = W_{tf_{t,d}} \times Idf_t \quad (1)$$

Keterangan:

$W_{tf_{t,d}}$: nilai *Term Frequency*.

Idf_t : nilai *Invers Document Frequency*.

Dengan nilai *IDF* dapat diperoleh dari Persamaan 2.

$$Idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right) \quad (2)$$

Keterangan:

N : jumlah total semua dokumen teks yang ada.

$df_{(t)}$: jumlah total seberapa sering munculnya pada *term*.

Tahap selanjutnya adalah penyusunan model dengan menggunakan metode *Naive Bayes Classifier*. Persamaan dari teorema bayes dirumuskan menggunakan Persamaan 3.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (3)$$

keterangan :

x : Data dengan kelas yang masih belum diketahui

c : Hipotesis data suatu kelas spesifik

$P(c|x)$: Nilai *posterior probability*

$P(c)$: Nilai *prior probability*

$P(x|c)$: Nilai *Conditional Probability*

$P(x)$: Kemungkinan c

Rumus perhitungan *prior probability* dirumuskan menggunakan Persamaan 4.

$$P(V_j) = \frac{D_j}{D} \quad (4)$$

Keterangan :

$P(V_j)$: *prior probability* pada suatu kelas

D : jumlah total dokumen yang ada

D_j : jumlah dokumen pada suatu kelas j

Rumus *conditional probability* dapat dinyatakan dalam Persamaan 5.

$$P(X_1|V_j) = \frac{W_{i,j}+1}{N+N_j} \quad (5)$$

Keterangan :

$P(X_1|V_j)$: *Conditional probability* pada kata ke- n dengan diketahuinya kelas j

$W_{i,j}$: Nilai pembobotan kata *TF-IDF* atau W dari term i di kelas j

- N : Jumlah kata unik pada semua dokumen yang ada
 - N_j : Jumlah total N dari seluruh term yang ada didalam kelas j
- Rumus perhitungan untuk hasil prediksi dapat dinyatakan sebagai Persamaan 6.

$$V_{map} = \prod_{i=1}^n P(X_i | V_j) P(V_j) \tag{6}$$

Setelah seluruh proses diatas ini telah dilakukan, dilanjutkan dengan melakukan uji evaluasi 4 jenis model dengan perbandingan data training dan data testing yang berbeda yaitu 60:40%, 70:30%, 80:20%, dan 90:10% dengan *confusion matrix* untuk mengetahui nilai akurasi, presisi, *recall*, dan *f-measure* masing-masing model dan mengetahui model terbaik dengan nilai akurasi yang paling tinggi. Terakhir yaitu melakukan uji sentimen dengan menggunakan model yang terbaik pada *dataset* yang berbeda.

3. Hasil dan Pembahasan

3.1. Hasil Data Mentah

Data mentah pada penelitian ini didapatkan dari proses *crawling data tweet* pada Twitter dengan menggunakan kode pemrograman *python* yang dijalankan di Google Collabs dan dengan bantuan Twitter API dari akun Twitter developer milik penulis. Proses *crawling data* dilakukan dengan mencari *tweet* dengan 10 kata kunci paling trending sepanjang 2021. Jumlah data yang didapatkan adalah sebanyak 2000 data *tweet*. Seluruh data mentah yang telah dikumpulkan akan dilabeli oleh Ahli Bahasa yaitu seorang Sarjana Jurusan Bahasa dan Sastra Indonesia. Pelabelan dilakukan dengan menggunakan teori *Natural Processing Language* berbasis konteks yang berarti memahami isi dan konteks dari setiap data lalu menentukan sentimennya. Kelompok sentimen di dalam penelitian ini ada 3 yaitu positif, negatif, dan netral. Contoh hasil pelabelan data mentah dapat dilihat pada Tabel 1.

Tabel 1. Contoh Pelabelan data mentah

Data <i>tweet</i>	Sentimen
Berharap gak usah pake vaksin lah, ribet, gak melindungi dari tertular dan menularkan juga.	Negatif
Vaksin atau gak vaksin, gak ada urusannya sama Aqidah. Aneh banget ampe bawa dalil untuk membenaran vaksin atau menolak vaksin.	Netral
Ayo #VaksinasiCovid19 😊 Jangan kira dirimu kebal virus	Positif

3.2. Hasil Pembuatan Kamus *Stopword Remover*

Penyusunan kamus *stopword remover* daperoleh dari hasil analisa penulis dengan bantuan dari ahli bahasa dengan cara memilih kata-kata yang merupakan kesalahan ketik dan kata yang kurang atau tidak jelas artinya di dalam isi *tweet* pada saat proses pelabelan *data tweet* dari hasil pengumpulan *dataset*. Hasil pembuatan Kamus *Stopword Remover* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pembuatan Kamus *Stopword Remover*

['m', 'ha', 'toh', 'huuu', 'mah', 'ko', 't', 'k', 'xx', 'da', 'tossss', 'dn', 'hm', 'o', 'akshsksk', 'x', 'ah', 'aaaa', 'njem', 'yyyea', 'nihh', 'oh', 'owalah', 'heh', 'huhu', 'hhdyeyegg', 'shshsh', 'lol', 'eh', 'nder', 'yang', 'di', 'dan', 'atau', 'dari', 'ke', 'dong', 'kepada', 'ikfina', 'nya', 'dki', 'ihh', 'klmj', 'kspi', 'yanc', 'bebb', 'nyaa', 'hzn', 'hira', 'mung', 'bgdddd', 'dftr', 'bossss', 'ylki', 'lpbi', 'gess', 'lek', 'lha', 'wehh', 'dll', 'garr', 'mmm', 'hrt', 'mabeeyah', 'ri', 'ih', 'jt', 'lah', 'iseh', 'nok', 'ueueue', 'akn', 'kog', 'kwn', 'dh', 'xixixi', 'gt', 'moel', 'wes', 'sep', 'hlo', 'aaaa', 'degdeg', 'yb', 'grgr', 'nai', 'plis', 'tkf', 'koq', 'wong', 'sy', 'yyyyy', 'ky', 'nggg', 'muah', 'aaa', 'sm', 'tun', 'yash', 'wah', 'kaa', 'unt', 'nga', 'guys', 'xé', 'too', 'coy', 'jir', 'az', 'geyss', 'sik', 'fpks', 'ôî', 'tbtb', 'pnh', 'by', 'yak', 'hmm', 'gaess', 'nyok', 'olll', 'ô', 'aw', 'jt', 'oooo']

3.3. Hasil *Pre-processing*

Tahapan *pre-processing* diawali dengan proses *cleansing*, yaitu dengan melakukan beberapa hal seperti penggantian *emoticon* pada data *tweet* yang ada dengan text yang menjelaskan ekspresi dari *emoticon*

tersebut. *Cleansing* juga akan menghapus beberapa hal seperti angka, simbol, link url, dan simbol hashtag. Hasil dari tahap *cleansing* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Cleansing*

Data mentah	<i>Cleansing</i>
Berharap gak usah pake vaksin lah, ribet, gak melindungi dari tertular dan menularkan juga. Vaksin atau gak vaksin, gak ada urusannya sama Aqidah. Aneh banget ampe bawa dalil untuk pembenaran vaksin atau menolak vaksin. Ayo #VaksinasiCovid19 😊 Jangan kira dirimu kebal virus	Berharap gak usah pake vaksin lah ribet gak melindungi dari tertular dan menularkan juga Vaksin atau gak vaksin gak ada urusannya sama Aqidah Aneh banget ampe bawa dalil untuk pembenaran vaksin atau menolak vaksin Ayo Vaksinasicovid grinningfacewithbigeyes Jangan kira diri kebal virus

Setelah melewati proses *cleansing*, selanjutnya ialah melakukan proses *casefolding* yaitu mengganti semua huruf yang ada pada kalimat di dalam *dataset* menjadi huruf kecil / *lowercase*. Hasil dari *dataset* yang telah melewati proses *casefolding* dapat dilihat pada Tabel 4.

Tabel 4 Hasil *Case Folding*

<i>Cleansing</i>	<i>Case Folding</i>
Berharap gak usah pake vaksin lah ribet gak melindungi dari tertular dan menularkan juga Vaksin atau gak vaksin gak ada urusannya sama Aqidah Aneh banget ampe bawa dalil untuk pembenaran vaksin atau menolak vaksin Ayo Vaksinasicovid grinningfacewithbigeyes Jangan kira diri kebal virus	berharap gak usah pake vaksin lah ribet gak melindungi dari tertular dan menularkan juga vaksin atau gak vaksin gak ada urusannya sama aqidah aneh banget ampe bawa dalil untuk pembenaran vaksin atau menolak vaksin ayo vaksinasicovid grinningfacewithbigeyes jangan kira diri kebal virus

Setelah melakukan proses *case folding*, maka selanjutnya adalah proses *tokenizing* atau melakukan pemisahan kalimat menjadi kata per kata yang nantinya tiap kata tersebut akan dikenal sebagai *token*. *Token* akan dipisah berdasarkan spasi yang ada di dalam kalimat pada *dataset*. Hasil dari *dataset* yang telah melewati proses *tokenizing* dapat dilihat pada Tabel 5

Tabel 5. Hasil *Tokenizing*

<i>Case Folding</i>	<i>Tokenizing</i>
berharap gak usah pake vaksin lah ribet gak melindungi dari tertular dan menularkan juga	['berharap', 'gak', 'usah', 'pake', 'vaksin', 'lah', 'ribet', 'gak', 'melindungi', 'dari', 'tertular', 'dan', 'menularkan', 'juga']
vaksin atau gak vaksin gak ada urusannya sama aqidah aneh banget ampe bawa dalil untuk pembenaran vaksin atau menolak vaksin	['vaksin', 'atau', 'gak', 'vaksin', 'gak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'ampe', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'atau', 'menolak', 'vaksin']
ayo vaksinasicovid grinningfacewithbigeyes jangan kira diri kebal virus	['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']

Lalu tahap setelah *tokenizing* adalah proses *normalization* atau melakukan penggantian kata-kata yang tidak baku atau salah ketik menjadi kata baku dengan tujuan agar dapat dimengerti makna dari kata-kata tersebut. Kamus perbaikan kata atau *normalization* yang digunakan didapatkan dari penelitian oleh Farah Dhaifa [10] dengan tambahan kata-kata dari peneliti dan ahli bahasa. Hasil dari *dataset* yang telah melewati proses *normalization* dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Normalization*

<i>Tokenizing</i>	<i>Normalization</i>
['berharap', 'gak', 'usah', 'pake', 'vaksin', 'lah', 'ribet', 'gak', 'melindungi', 'dari', 'tertular', 'dan', 'menularkan', 'juga']	['berharap', 'tidak', 'perlu', 'pakai', 'vaksin', 'lah', 'ribet', 'tidak', 'melindungi', 'dari', 'tertular', 'dan', 'menularkan', 'juga']
['vaksin', 'atau', 'gak', 'vaksin', 'gak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'ampe', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'atau', 'menolak', 'vaksin']	['vaksin', 'atau', 'tidak', 'vaksin', 'tidak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'atau', 'menolak', 'vaksin']
['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']	['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']

Setelah melewati tahap *normalization*, selanjutnya ialah melakukan proses *stopword remover*. Pada proses ini akan dilakukan penghapusan kata-kata yang tidak memiliki makna atau arti penting didalamnya dengan tujuan untuk mempercepat pemrosesan *dataset* dengan hanya mengambil kata-kata yang bermakna penting saja. Hasil dari *dataset* yang telah melewati proses *stopword remover* dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Stopword Remover*

<i>Normalization</i>	<i>Stopword Remover</i>
['berharap', 'tidak', 'perlu', 'pakai', 'vaksin', 'lah', 'ribet', 'tidak', 'melindungi', 'dari', 'tertular', 'dan', 'menularkan', 'juga']	['berharap', 'tidak', 'perlu', 'pakai', 'vaksin', 'ribet', 'tidak', 'melindungi', 'tertular', 'menularkan', 'juga']
['vaksin', 'atau', 'tidak', 'vaksin', 'tidak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'atau', 'menolak', 'vaksin']	['vaksin', 'tidak', 'vaksin', 'tidak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'menolak', 'vaksin']
['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']	['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']

Setelah menyelesaikan proses *stopword removal*, selanjutnya ialah melakukan proses *stemming* atau menghapus imbuhan awalan atau akhiran dari setiap kata yang ber-imbuhan ada sehingga membuat semua kata-kata yang ada di dalam *dataset* menjadi kata dasar. Hasil dari *dataset* yang telah melewati proses *stemming* dapat dilihat pada Tabel 8.

Tabel 8. Hasil *Stemming*

<i>Normalization</i>	<i>Stopword Remover</i>
['berharap', 'tidak', 'perlu', 'pakai', 'vaksin', 'ribet', 'tidak', 'melindungi', 'tertular', 'menularkan', 'juga']	['harap', 'tidak', 'perlu', 'pakai', 'vaksin', 'ribet', 'tidak', 'lindung', 'tular', 'tular', 'juga']
['vaksin', 'tidak', 'vaksin', 'tidak', 'ada', 'urusannya', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'pembenaran', 'vaksin', 'menolak', 'vaksin']	['vaksin', 'tidak', 'vaksin', 'tidak', 'ada', 'urus', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'benar', 'vaksin', 'tolak', 'vaksin']
['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']	['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']

3.4. Hasil Pembobotan Kata

Pembobotan kata di dalam penelitian ini adalah implementasi algoritma *Term Frequency-Inverse Document Frequency (TF-IDF)*. Algoritmi ini dimulai dari penentuan nilai *Term Frequency (TF)* dengan cara menghitung frekuensi seberapa banyak munculnya sebuah kata dari satu kalimat tekstual dengan mengimplementasikan kode program *feature_extraction* dari *TfidfVectorizer* pada *library scikit-learn python*. Contoh implementasi perhitungan *Term Frequency* pada *dataset* yang telah menyelesaikan tahap *preprocessing* dapat dilihat pada Tabel 9.

Tabel 9. Hasil perhitungan *TF*

<i>Normalization</i>	<i>Term Frequency (TF)</i>
----------------------	----------------------------

['harap', 'tidak', 'perlu', 'pakai', 'vaksin', 'ribet', 'tidak', 'lindung', 'tular', 'tular', 'juga']	Kata	Nilai TF
	harap	1/11
	tidak	2/11
	perlu	1/11
	pakai	1/11
	vaksin	1/11
	ribet	1/11
	lindung	1/11
	tular	2/11
juga	1/11	
['vaksin', 'tidak', 'vaksin', 'tidak', 'ada', 'urus', 'sama', 'aqidah', 'aneh', 'banget', 'sampai', 'bawa', 'dalil', 'untuk', 'benar', 'vaksin', 'tolak', 'vaksin']	Kata	Nilai TF
	vaksin	4/18
	tidak	2/18
	ada	1/18
	urus	1/18
	sama	1/18
	aqidah	1/18
	aneh	1/18
	banget	1/18
	sampai	1/18
	bawa	1/18
	dalil	1/18
	untuk	1/18
	benar	1/18
tolak	1/18	
['ayo', 'vaksinasicovid', 'grinningfacewithbigeyes', 'jangan', 'kira', 'diri', 'kebal', 'virus']	Kata	Nilai TF
	ayo	1/8
	vaksinasicovid	1/8
	grinningfacewithbigeyes	1/8
	jangan	1/8
	kira	1/8
	diri	1/8
	kebal	1/8
virus	1/8	

Setelah perhitungan *Term Frequency (TF)* telah selesai dan nilai TF setiap kata telah ditentukan, selanjutnya adalah melaksanakan proses perhitungan untuk menentukan *Document Frequency (DF)* dan *Invers Document Frequency (IDF)*. Dalam tahap ini dilakukan proses perhitungan nilai *DF* setiap kata terlebih dahulu. Setelah nilai *DF* telah berhasil ditentukan, selanjutnya dilakukan perhitungan nilai *IDF* setiap kata dengan cara menghitung nilai log dari nilai *DF*. Setelah nilai *Document Frequency (DF)* dan *Invers Document Frequency (IDF)* setiap kata telah ditentukan, maka tahap selanjutnya ialah menghitung nilai *Term Frequency – Invers Document Frequency (TF-IDF)* dengan cara melakukan perkalian antara *Term Frequency (TF)* dan *Invers Document Frequency (IDF)*. Contoh hasil nilai *DF* dan *IDF* beserta hasil akhir *TF-IDF* setiap kata yang ada di dalam data dapat dilihat pada Tabel 10.

Tabel 10. Hasil Perhitungan *DF* dan *IDF* beserta hasil akhir *TF-IDF*

No	Kata	Nilai <i>DF</i>	$Idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right)$	$W_{t,d} = Wtf_{t,d} \times Idf_t$
1	harap	13	2.1870866433	0.1988260585
	tidak	328	0.7851561519	0.1427556640
	perlu	40	1.6989700043	0.1544518186
	pakai	40	1.6989700043	0.15445181857

	vaksin	1720	0.0655015487	0.00595468625
	ribet	2	3.0	0.27272727273
	lindung	149	1.1278437272	0.10253124793
	tular	48	1.6197887582	0.29450704695
	juga	95	1.3233063903	0.14725352347
2	vaksin	1720	0.0655015487	0.01455589971
	tidak	328	0.7851561519	0.08723957243
	ada	172	1.0655015488	0.05919453048
	urus	14	2.1549019599	0.11971677555
	sama	102	1.2924298239	0.07180165688
	aqidah	1	3.3010299956	0.18339055531
	aneh	8	2.3979400086	0.13321888936
	banget	93	1.3325470471	0.07403039151
	sampai	60	1.5228787453	0.08460437473
	bawa	6	2.5228787453	0.14015993029
	dalil	1	3.3010299956	0.18339055531
	untuk	205	0.9892761346	0.05495978525
	benar	35	1.7569619513	0.09760899729
	tolak	21	1.9788107009	0.10993392783
3	ayo	580	0.5376020021	0.06720025026
	vaksinasicovid	55	1.5606673061	0.19508341326
	grinningfacewithbigeyes	1	3.3010299956	0.41262874945
	jangan	295	0.8312079796	0.10390099745
	kira	3	2.8239087409	0.35298859261
	diri	99	1.3053948011	0.16317435013
	kebal	24	1.9208187540	0.24010234425
	virus	85	1.3716110699	0.17145138373

3.5. Penyusunan dan Hasil Pengujian Model Naive Bayes Classifier

Penerapan metode *Naive Bayes Classifier* pada tahap penyusunan model dengan cara *training data* yang dimulai dengan penentuan nilai *prior probability* cara dengan membagi jumlah dokumen dalam suatu kelas tertentu dengan jumlah total dokumen yang ada. hasil penentuan nilai *prior probability* dapat dilihat pada Tabel 11.

Tabel 11. Hasil *Prior Probability*

Sentimen	$P(V_j) = \frac{D_j}{D}$
Negatif	0.13875
Netral	0.071875
Positif	0.789375

Hasil nilai *prior probability* setiap kelas akan digunakan untuk melakukan proses penentuan hasil perhitungan suatu prediksi. Setelah proses penentuan nilai *prior probability* telah selesai, maka dilanjutkan dengan penentuan nilai *conditional probability* atau penentuan bobot kata terhadap semua kelas prediksi. Contoh hasil penentuan nilai *conditional probability* dapat dilihat pada Tabel 12.

Tabel 12. Hasil *Conditional Probability*

Sentimen	Kata	$P(X_1 V_j) = \frac{W_{i,j} + 1}{N + N_j}$	$P(X_1 V_j) = \frac{W_{i,j} + 1}{N + N_j}$	$P(X_1 V_j) = \frac{W_{i,j} + 1}{N + N_j}$
		Positif	Netral	Negatif
Negatif	harap	0.00044058289	0.00044220806	0.00044171925
	tidak	0.00039637726	0.00041047258	0.00040885712
	perlu	0.00042210304	0.00042458691	0.00042458691
	pakai	0.00042179459	0.00042505589	0.00042427483
	vaksin	0.00024943086	0.00035358688	0.00033813602
	ribet	0.00046946782	0.00046946782	0.00046894888
	lindung	0.00038563527	0.00040653807	0.00040593934

Netral	tular	0.00047004613	0.00047750167	0.00047662262
	juga	0.00041282962	0.00042101046	0.00042101046
	vaksin	0.00023710116	0.00035573488	0.00034011260
	tidak	0.00037712091	0.00039053145	0.00038899448
	ada	0.00036790362	0.00038474192	0.00038004827
	urus	0.00041241870	0.00041211511	0.00041196349
	sama	0.00038402065	0.00039389991	0.00039361059
	aqidah	0.00043651440	0.00043635344	0.00043651440
	aneh	0.00041739185	0.00041785357	0.00041754564
	banget	0.00038928249	0.00039515466	0.00039069857
	sampai	0.00039411496	0.00039889826	0.00039845862
	bawa	0.00041994840	0.00042025799	0.00042056803
	dalil	0.00043651440	0.00043635344	0.00043651440
	untuk	0.00036643271	0.00038685727	0.00038614926
Positif	benar	0.00040279229	0.00040353271	0.00040308813
	tolak	0.00040836421	0.00040851451	0.00040821402
	ayo	0.00032300249	0.00039322043	0.00039264174
	vaksinasicovid	0.00043315817	0.00044017805	0.00044034024
	grinningfacew	0.00052088080	0.00052107294	0.00052107294
	ithbigeyes			
	jangan	0.00037496637	0.00040200327	0.00040258971
	kira	0.00049870571	0.00049907362	0.00049888959
	diri	0.00041616255	0.00042795229	0.00042779490
	kebal	0.00045341950	0.00045743354	0.00045743354
virus	0.00041957427	0.00043163278	0.00043195110	

Setelah tahap penentuan nilai *conditional probability* telah selesai dan telah didapatkan nilai per katanya, maka selanjutnya adalah melaksanakan tahap *testing*. Dalam tahapan ini metode *Naive Bayes Classifier* memilih nilai probabilitas yang paling tinggi diantara ketiga kelas target dengan menggunakan rangkaian kode program *scikit-learn library python*. Contoh hasil *testing* setiap data dapat dilihat pada Tabel 13.

Tabel 13. Hasil Testing data

Data (Sentimen Prediksi)	V_{map} Positif	V_{map} Netral	V_{map} Negatif	Kelas Target
(NEGATIF) Berharap gak usah pake vaksin lah, ribet, gak melindungi dari tertular dan menularkan juga.	0.00032422	0.00002275	0.00080089	Negatif
(NETRAL) Vaksin atau gak vaksin, gak ada urusannya sama Aqidah. Aneh banget ampe bawa dalil untuk membenaran vaksin atau menolak vaksin.	0.00058688	0.00286371	0.00282657	Netral
(POSITIF) Ayo #VaksinasiCovid19 😊 Jangan kira dirimu kebal virus	0,00014820	0,00002100	0,00002100	Positif

3.6. Hasil Uji Evaluasi

Tahapan uji evaluasi pada penelitian ini menggunakan *confusion matrix* dengan tujuan untuk mengetahui seberapa baik tingkat akurasi pada proses klasifikasi yang telah diproses oleh sistem dengan menggunakan acuan nilai akurasi. Dalam penelitian ini, tahapan ini dilakukan sebanyak empat kali percobaan dengan menggunakan *data training* dan *data testing* masing-masing sebanyak 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Dalam pengimplementasiannya, perhitungannya akurasi, presisi, *recall*, dan *f-measure* juga digunakan dalam penelitian ini dengan menggunakan rangkaian kode program *library scikit-learn python*. Hasil data uji evaluasi pada tahapan ini dapat dilihat pada Tabel 14.

Tabel 14. Hasil Uji Evaluasi

Keterangan	Percobaan 1	Percobaan 2	Percobaan 3	Percobaan 4
Data Train	60%	70%	80%	90%
Data Testing	40%	30%	20%	10%
True Positif (TP)	596	444	295	151
True Negatif (TN)	52	38	26	15
True Netral (TL)	20	15	10	6
False Positif Data Netral (FP)	25	16	10	1
False Positif Data Negatif (FP)	40	28	17	7
False Negatif Data Positif (FN)	18	17	14	4
False Negatif Data Netral (FN)	12	12	9	7
False Netral Data Positif (FL)	18	13	7	3
False Netral Data Negatif (FL)	19	17	12	6
Akurasi	83.5	82.83	82.75	86
Presisi	0.823	0.821	0.824	0.859
Recall	0.833	0.831	0.823	0.864
F-measure	82.67	82.17	82.02	85.76

3.7. Hasil Uji Sentimen

Kegiatan evaluasi uji sentimen perlu dilakukan dengan tujuan untuk dapat mengetahui sebuah data *tweet* yang diuji apakah memiliki sentimen negatif atau positif atau netral. Pada tahapan ini, data *tweet* yang dipakai adalah data yang diperoleh dari hasil kegiatan *crawling data* diluar tanggal pengambilan *dataset* awal yaitu pada bulan Desember 2021. Total data mentah baru yang didapatkan adalah 10 data. Proses *crawling data* dilakukan dengan cara dan tahapan yang sama dengan *crawling data* awal. Setelah data mentah berhasil terkumpul maka dilanjutkan dengan melakukan proses pelabelan data dengan bantuan Ahli Bahasa. Setelah proses pelabelan data selesai, maka dilanjutkan dengan memulai proses uji sentimen. Proses pengujian ini dilaksanakan dengan menggunakan model yang memiliki akurasi terbaik diantara 4 percobaan yang sudah dilakukan sebelumnya yaitu porsi perbandingan antara *data training* dan *data testing* sebesar 90%:10%. Hasil dari seluruh percobaan pada keseluruhan tahapan uji sentimen ini dapat dilihat pada Tabel 15.

Tabel 15. Hasil Uji Sentimen

Data Tweet	Hasil Uji Kelas Sentimen	Sentimen yang dilabeli Ahli Bahasa
Yuk vaksin guys, kalo bukaan kita yg menjaga kesehatan siapa lagi dong #SukseskanVaksinCovid #VaksinasiCovid-19	Positif	Positif
Yg blom vaksin ayo cepetan vaksin ya guyysss, demi kesehatan bersama #VaksinasiCovid #AyoVaksin	Positif	Positif
Mumpung vaksin sekarang masih gratis, ayo segera antre vaksin sebelum kehabisan #YukAntreVaksin	Positif	Positif
Dah jgn ikutan vaksin, Pemerintah cuma pengen berbisnis dg rakyatnya aja. Sealu saja jadi ajang buat cari duitt #StopJualanVaksin	Negatif	Negatif
Kanapa sih pas mau vaksin slalu ae habis stoknya padaahal udh antre dari pagi tadi tapi tetep aja gak kebagian	Negatif	Negatif
Alhamdulillah udh ikutan vaksin kedua. Yuk yg blum vaksin segera ikutan vaksin yaaa, biar segera terwujud herd immunity ini #VaksinGratis #YukVaksin #SukseskanVaksin	Positif	Positif
Itu ya terserah dia lah mau ngikut vaksin atau tidak. Itu kan haknya dia jdi kita ga boleh menghakimi keputusannya apapun itu #TolakPerpecahan	Netral	Netral
Gw ga mau vaksin karena manfaatnya aja masih kurang yakin gw. Trus abis vaksin juga msih disuruh jaga jarak n dibatasi mau kemana mana ya buat apa coba ikutan vaksin #TolakVaksin	Negatif	Negatif
Ayoo sukseskan program vaksin demi kesehatan kita n keluarga kita di rumah. Kalo klian syg ama keluarga kalian, harusnya klian ikutan vaksin dong #VaksinMelindungi	Positif	Positif

Tolong yaa. Yg percaya vaksin jgn ngejek yg ga percaya vaksin. Smua org punya pilihannya masing masing, yg ga percaya vaksin juga jangan nyebarin berita hoax #StopHoax #StopPerpecahan




Positif

Netral

Keterangan: Sentimen yang berwarna merah menandakan bahwa perbandingan hasil uji prediksi berbeda dengan sentimen yang dilabeli oleh ahli bahasa.

Berdasarkan data dan proses yang telah dilakukan pada penelitian ini dalam file *excel*, diketahui bahwa dari 2000 data *tweet* dari tulisan masyarakat tentang opini mereka mengenai program *vaksinasi* yang ada di Indonesia pada *twitter* yang telah dikumpulkan sebelumnya 78% atau 1579 diantaranya adalah data yang bersifat positif, 13% atau 277 diantaranya adalah data yang bersifat negatif, dan 8% atau 144 diantaranya adalah data yang bersifat netral. Dalam 2000 *tweet* tersebut, Setiap data *tweet* memiliki beragam kata dan juga mengandung arti yang berbeda-beda. Seluruh data yang berjumlah total 2000 *tweet* disimpan di dalam sebuah tabel *excel* yang memiliki 2000 baris yang setiap barisnya berisi satu data *tweet*. Berikutnya dilakukan konversi data *tweet* dari tiap sentimen menjadi terpisah-pisah per kata untuk nantinya dimasukkan kedalam *word-cloud* menggunakan tools *word-cloud generator* dengan tujuan untuk mengetahui rata-rata isi *tweet* dari setiap sentimen. Berikut ini dapat dilihat hasil *word-cloud* dari setiap sentimen dan rata-rata isi kandungan *tweet*-nya pada Tabel 16.

Tabel 16. *Word-cloud* dan isi kandungan rata-rata setiap sentimen

<i>Word-cloud</i>	Rata-rata isi kandungan <i>tweet</i>
<p>Sentimen Positif</p> 	<ul style="list-style-type: none"> Dukungan terhadap program <i>vaksinasi</i>, Kepercayaan masyarakat terhadap suksesnya program ini, Memotivasi masyarakat lainnya untuk segera mengikuti program ini agar penyebarannya dapat dihentikan, Dukungan dan himbauan agar lebih peduli terhadap kesehatan masing-masing, Setuju dengan program vaksin gratis, bersyukur dengan adanya program ini.
<p>Sentimen Negatif</p> 	<ul style="list-style-type: none"> Ketakutan masyarakat terhadap bahan kimia yang ada didalam vaksin, Masyarakat merasa kesusahan dalam mengantri vaksin, Tidak setuju dengan vaksin berbayar atau vaksin yang dijual, Penolakan terhadap program vaksin, Keluhan masyarakat yang masyarakat tidak kebagian vaksin, Syarat vaksin yang masih susah
<p>Sentimen Netral</p> 	<ul style="list-style-type: none"> Masyarakat yang kebingungan dan ingin bertanya mengenai vaksin, Masyarakat yang menolak keributan antara pihak yang mendukung dan menolak program vaksin, Masyarakat yang mengajak untuk jangan percaya dengan berita hoax, Masyarakat yang setuju dengan adanya program vaksin tetapi menurut mereka masih ada yang perlu dibenahi didalamnya

Disimpulkan bahwa 2000 data opini publik pada Twitter terhadap program vaksinasi *covid-19* di Indonesia pada rentang waktu 8 Agustus sampai dengan 1 Oktober 2021 menunjukkan bahwa jumlah hasil data sentimen positif lebih dominan ke arah sentimen positif dengan jumlah data sebanyak 78% daripada sentimen negatif yang jumlah datanya sebanyak 13% maupun netral yang jumlah datanya sebanyak 7%. Uji sentimen menunjukkan hasil sentimen negatif apabila data *tweet* yang diuji memiliki kata-kata yang bermakna negatif seperti penolakan, keraguan, dan pengecaman. Sedangkan uji sentimen menunjukkan hasil sentimen positif apabila data *tweet* yang diuji memiliki kata-kata yang bermakna positif seperti pernyataan setuju, ajakan, pujian, dan dukungan. Lalu uji sentimen yang menunjukkan hasil sentimen netral apabila data *tweet* yang diuji tidak memiliki kata-kata yang bermakna negatif atau positif dan berisi kata-kata yang netral seperti setuju dengan perbedaan pendapat, dan menolak perpecahan atau ejekan akibat perbedaan pendapat.

4. Kesimpulan

Penggunaan metode *Naive Bayes Classifier* didalam penelitian ini bekerja dengan baik. Berdasarkan hasil pengujian didapatkan hasil akurasi yang berbeda pada keempat percobaan dengan rata-rata nilai akurasi sekitar 80-an %. Nilai akurasi tertinggi didapatkan pada percobaan keempat yakni sebesar 86% dengan data *training* 90% dan data *testing* 10%.

Daftar Pustaka

- [1] Widiyani, Rosmha. (2020). Latar Belakang Virus Corona, Perkembangan hingga Isu Terkini. <https://news.detik.com/berita/d-4943950/latar-belakang-virus-corona-perkembangan-hingga-isu-terkini> (Accessed July 13, 2021).
- [2] Nareza, Meva. (2021). Kenali Perbedaan Vaksin-Vaksin COVID-19 yang Akan Digunakan di Indonesia. <https://www.alodokter.com/kenali-perbedaan-vaksin-vaksin-covid-19-yang-akan-digunakan-di-indonesia> (Accessed July 13, 2021).
- [3] Syabriansyah, Rijal. (2021). Pro dan Kontra Vaksin COVID-19. <https://kumparan.com/rijal-syabriansyah-219/pro-dan-kontra-vaksin-covid-19-1vcXO9pvtz1> (Accessed July 13, 2021).
- [4] Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11 (2), 195–200. <https://doi.org/10.32736/sisfokom.v11i2.1412> (Accessed Oct. 16, 2022).
- [5] Wongkar, M., & Angdresy, A. (2019, October 1). Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter. Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019. <https://doi.org/10.1109/ICIC47613.2019.8985884> (Accessed Oct. 16, 2022).
- [6] Rozi, I. F., Pramono, S. H., & Dahlan E. A. 2013. Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. Malang. *Jurnal EECCIS*. 6(1): (Accessed July 13, 2021).
- [7] Pang, B., & Lee, L.(2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. <http://dx.doi.org/10.1561/1500000011> (Accessed July 13, 2021).
- [8] Alter Lasarudin, Purwanto. 2018. Klasifikasi Pengaduan Masyarakat Menggunakan Naive Bayes Berbasis Seleksi Atribut Informasion Gain. (Accessed July 13, 2021).
- [9] K. Verspoor and K. B. Cohen, "Natural Language Processing," in *Encyclopedia of Systems Biology*, New York, NY: Springer New York, 2013, pp. 1495–1498. doi: 10.1007/978-1-4419-9863-7_158.
- [10] Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis Sentimen kebijakan kampus merdeka menggunakan naive Bayes Dan Pembobotan TF-IDF Berdasarkan Komentar Pada YouTube. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 2(1). <https://doi.org/10.25126/justsi.v2i1.24> (Accessed Oct. 16, 2022).