

# Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Vaksinasi Booster 2 Menggunakan Metode *Naïve Bayes* *Classifier*

Alma Syahrir<sup>a,1</sup>, Harlinda<sup>a,2</sup>, Fitriyani Umar<sup>a,3</sup>

<sup>a</sup> Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM. 5, Makassar dan 90231, Indonesia  
<sup>1</sup> almaryahrir@gmail.com; <sup>2</sup> harlinda@umi.ac.id; <sup>3</sup> fitriyani.umar@umi.ac.id  
\*corresponding author

## INFORMASI ARTIKEL

Diterima : 26 – 07 – 2023  
Direvisi : 01 – 11 – 2023  
Diterbitkan : 04 – 12 – 2023

**Kata Kunci:**  
Accuracy  
Naïve Bayes Classifier  
Precision  
Recall  
Sentimen

## ABSTRAK

Upaya yang dilakukan pemerintah salah satunya yaitu mewajibkan seluruh masyarakat untuk melakukan vaksinasi sebanyak tiga kali yaitu vaksinasi pertama, kedua, dan ketiga atau biasa disebut dengan vaksin booster. Namun dengan adanya kebijakan pemerintah terkait pelaksanaan vaksinasi booster yang terjadi di masyarakat khususnya vaksinasi booster dua yang merupakan vaksinasi lanjutan banyak menuai pro dan kontra ditengah masyarakat. Beragam pendapat dari masyarakat ada yang mendukung program ini dan sebagian yang lainnya menolak dengan berbagai alasan. Oleh karena itu, salah satu cara untuk mengetahui tanggapan masyarakat terkait kebijakan ini adalah melalui analisis sentimen. Analisis sentimen merupakan pengamatan yang dikelola berdasarkan pendapat, sentimen, evaluasi, sikap, dan perasaan orang lain terutama berdasarkan apa yang mereka tulis. Pendapat lain memaparkan bahwa analisis sentimen bertujuan untuk mengetahui dan memprediksi apa yang difikirkan seseorang berdasarkan informasi seperti opini baik bersifat negatif maupun positif. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat terhadap kebijakan pemerintah vaksinasi booster 2 menggunakan metode *Naïve Bayes Classifier*. Hasil penelitian menunjukkan hasil pengujian dengan menggunakan *validation* maka memperoleh nilai *performance* yaitu *accuracy* sebesar 86.84, *precision* sebesar 85.71% dan *recall* sebesar 90%.

This is an open access article under the [CC-BY-SA](#) license



## I. Pendahuluan

Pada akhir tahun 2019 di Wuhan, tepatnya di China telah muncul wabah penyakit baru yang disebabkan oleh virus yang disebut dengan COVID-19. *World Health Organization* (WHO) menduga virus ini berasal dari hwan yang tertular kepada manusia, sifat dari virus tersebut dapat melalui kontak fisik secara langsung akibatnya penyebaran virus ini sangat cepat sehingga dengan mudah tersebar di seluruh Indonesia. Oleh karena itu, WHO menetapkan Covid-19 sebagai pandemi pada 9 Maret 2020. Di Indonesia, kasus virus corona pertama kali ditemukan di kota Depok yang dimana kasus ini segera di tindak langsung oleh pemerintah untuk menekan penyebaran kasus virus corona, sehingga berbagai upaya telah dilakukan oleh pemerintah salah satunya adalah vaksinasi [1].

Vaksinasi adalah salah satu cara untuk mengurangi bertambahnya angka kasus virus corona. Vaksin berfungsi untuk meningkatkan kekebalan tubuh seseorang agar dapat mencegah penyebaran virus tertentu. Meskipun vaksin tidak melindungi 100% dari ancaman virus corona, namun vaksin bisa memperkecil kemungkinan seseorang terjangkit virus corona [2].

Upaya yang dilakukan pemerintah salah satunya yaitu mewajibkan seluruh masyarakat untuk melakukan vaksinasi sebanyak tiga kali yaitu vaksinasi pertama, kedua, dan ketiga atau biasa disebut dengan vaksin booster [3]. Namun dengan adanya kebijakan pemerintah terkait pelaksanaan vaksinasi booster yang terjadi di masyarakat khususnya vaksinasi booster dua yang merupakan vaksinasi lanjutan banyak menuai pro dan kontra ditengah masyarakat. Beragam pendapat dari masyarakat ada yang mendukung program ini dan sebagian yang

lainnya menolak dengan berbagai alasan. Oleh karena itu, salah satu cara untuk mengetahui tanggapan masyarakat terkait kebijakan ini adalah melalui analisis sentimen.

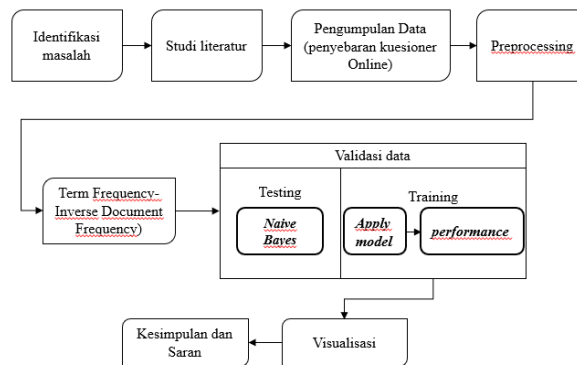
Analisis sentimen merupakan pengamatan yang dikelola berdasarkan pendapat, sentimen, evaluasi, sikap, dan perasaan orang lain terutama berdasarkan apa yang mereka tulis. Pendapat lain memaparkan bahwa analisis sentimen bertujuan untuk mengetahui dan memprediksi apa yang difikirkan seseorang berdasarkan informasi seperti opini baik bersifat negatif maupun positif [4].

Adapun penelitian terdahulu mengenai analisis sentimen yang telah dilakukan sebelumnya yaitu penelitian yang dilakukan pada tahun 2021 dimana peneliti menggunakan metode *Recurrent Neural Network* dan *Naive Bayes Classifier* dengan tujuan untuk membandingkan kinerja dari kedua metode tersebut. Dari hasil penelitian tersebut, diperoleh hasil akurasi dari metode RNN sebesar 97,77% dibandingkan dengan *Naive Bayes* dengan nilai akurasi sebesar 80% [5]. Penelitian selanjutnya pada tahun 2021 yang dimana penelitian ini bertujuan untuk menganalisis opini publik terhadap pembelajaran daring yang dilakukan dimasa pandemi di Indonesia. Penelitian yang dilakukan yaitu penambangan teks berbasis dokumen pada twitter yang dianalisis menggunakan algoritma *Naive Bayes*. Hasil dari penelitian tersebut menunjukkan bahwa pembelajaran daring sebanyak 30% yaitu sentimen positif, 69% sentimen negatif, dan 1% netral pada priode tersebut. Tingginya sentimen negatif tersebut dikarenakan ketidakpuasan masyarakat terhadap pembelajaran daring [6].

Pada penelitian terdahulu yang telah melakukan analisis sentimen yang berjudul “Analisis Sentimen Masyarakat di Twitter Terkait Pandemi Covid-19”, penelitian tersebut menggunakan metode *Naive Bayes* dimana pengujian dilakukan menggunakan *Cross Validation* meliputi 5 fold dan 10 fold. Hasil *Cross Validation* pengujian 5fold didapatkan hasil rata-rata sebesar 0.756364 (75%). Dan hasil pengujian 10 fold didapatkan rata-rata akurasi sebesar 0.76 (76%) [7]. Penelitian yang berjudul “Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode *Naive Bayes Classifier*”, dimana pada penelitian ini pengumpulan data dilakukan melalui kuesioner online dengan responden sebanyak 1070 yang berasal dari Universitas Islam Negeri Sumatra Utara Medan. Akurasi yang dihasilkan dari metode tersebut dapat menerapkan nilai akurasi yang cukup baik dengan rata-rata 83% [8]. Adapun Penelitian yang berjudul “Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 Pada Media Sosisal Twitter”, dimana penelitian ini bertujuan untuk menganalisis respon masyarakat terhadap wacana vaksinasi tahap pertama dengan mengelompokkan opini masyarakat menggunakan metode *Latent Dirichlet Allocation (LDA)*. Pada hasil analisis menunjukkan bahwa masyarakat lebih banyak memberikan respon positif (30%) dibandingkan dengan respon negatif (26%) [9]. Penelitian selanjutnya yang berjudul “Analisis Sentimen Jasa Transportasi Online Pada Twitter Menggunakan Metode *Naive Bayes Classifier*”, dimana hasil analisis menyatakan bahwa setelah menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. [10]. Adapun penelitian yang berjudul “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan *Naive Bayes Classifier*”, berdasarkan hasil penelitian tersebut diketahui bahwa opini masyarakat terkait kebijakan PPKM dihasilkan 99% klasifikasi polaritas positif dan 1% polaritas negatif [11]. Berdasarkan permasalahan tersebut dan latar belakang yang terjadi maka penulis mengangkat topik yang berjudul “Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Vaksinasi Booster 2 Menggunakan Metode *Naive Bayes Classifier*”. Dimana tujuan dari penelitian ini yaitu untuk mengetahui seberapa besar tanggapan positif maupun negatif dari masyarakat terhadap adanya kebijakan pemerintah terkait pelaksanaan program vaksinasi booster 2 di Kabupaten Pinrang priode 2023.

## II. Metode

Metode penelitian merupakan suatu gambaran atau langkah-langkah yang dilakukan dalam penelitian yang dimana metodologi penelitian tersebut bertujuan agar pelaksanaan penelitian dapat berjalan secara terstruktur.



Gambar 1. Tahapan Penelitian

### 1. Identifikasi Masalah

Pada tahap ini, terdapat kegiatan mengidentifikasi masalah dari permasalahan yang melatar belakangi penelitian yang akan dilakukan yaitu permasalahan mengenai pelaksanaan vaksinasi booster dua yang melahirkan pro dan kontra di tengah masyarakat.

### 2. Studi Literatur

Pada tahap ini, pencarian literatur berupa jurnal yang berkaitan dengan penelitian yang akan di bahas mengenai analisis sentimen masyarakat terhadap pelaksanaan vaksin di Indonesia. Literatur juga berfungsi sebagai referensi untuk membantu dalam proses penelitian ini.

### 3. Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini menggunakan teknik pengambilan data dengan melakukan penyebaran kuesioner secara online kepada masyarakat mengenai adanya pelaksanaan vaksinasi booster dua guna memperoleh informasi yang dapat mendukung penelitian ini.

### 4. Preprocessing

Selanjutnya pada tahap ini data mentah yang dikumpulkan akan diseleksi terlebih dahulu agar bisa digunakan untuk pemrosesan lebih lanjut, maka diperlukan *preprocessing* yang dilakukan dalam enam tahap yaitu *Cleaning*, *Case Folding Normalization*, *Tokenizing*, *Stemming*, dan *Stopword Removal*.

### 5. Pembobotan kata TF-IDF

Pada tahap ini dilakukan pembobotan kata TF-IDF untuk mengetahui nilai bobot dari setiap kata (*term*).

### 6. Validasi data

Selanjutnya pada proses analisis data terdiri dari pengklasifikasian menggunakan metode *Naive Bayes Classifier* dan dilanjutkan validasi data untuk menentukan nilai performance.

### 7. Visualisasi

Pada tahap ini, dilakukan visualisasi data untuk mempresentasikan hasil analisis agar lebih jelas dan mudah dipahami.

### 8. Kesimpulan dan saran

Pada tahap ini penulis memberikan kesimpulan terhadap hasil dari penelitian analisis sentimen yang telah dilakukan agar rumusan masalah sebelumnya dapat terjawab dan diberikan Saran agar dapat menjadi suatu pembelajaran baru untuk kedepannya.

#### a. *Naive Bayes Classifier*

Merupakan metode klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel yang diberikan termasuk dalam kelas tertentu. Metode ini sangat cocok digunakan sebagai pengklasifikasian sentimen dikarenakan memiliki metode yang sederhana, tepat, terbukti, memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Perhitungan kategori probabilitas dalam *Naive Bayes* menggunakan pendekatan algoritma *bayes* menggunakan persamaan [12][13] [14] :

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

Ket:

Y = Suatu kelas spesifik

X = Data pada kelas yang belum diketahui

$P(Y|X)$  = Probabilitas hipotesis Y didasarkan pada kondisi X

$P(Y)$  = Probabilitas Y

$P(X|Y)$  = Probabilitas hipotesis X didasarkan pada kondisi Y

$P(X)$  = Probabilitas X

b. Analisis Validasi Data

a) *Accuracy*

*Accuracy* merupakan perhitungan tingkat ketepatan dari suatu kelas prediksi untuk mengukur keakuratan keseluruhan model yang telah dibuat [10]. Rumus akurasi ditunjukkan dengan persamaan sebagai berikut:

$$akurasi = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (3)$$

b) *Precision*

*Precision* merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna terhadap jawaban yang dihasilkan. Adapun persamaannya sebagai berikut:

$$presisi = \frac{(TP)}{(TP+FP)} \quad (4)$$

c) *Recall*

*Recall* merupakan perhitungan kinerja yang memberikan sebuah informasi mengenai prediksi data positif yang diprediksi sebagai data negatif. Perhitungan ini dapat dilakukan menggunakan rumus sebagai berikut:

$$recall = \frac{(TP)}{(TP+FN)} \quad (5)$$

d) *F1-Score*

*F1-score* atau disebut dengan *f-measure* merupakan salah satu perhitungan evaluasi dalam data yang mengkombinasikan *recall* dan *precision* untuk mendapatkan nilai yang lebih akurat [15]. Rumus dari *f-measure* adalah sebagai berikut:

$$f - measure = \frac{2 \cdot recall \cdot precision}{recall + precision} \quad (6)$$

Keterangan:

TP : *True positive*

TN : *True negative*

FP : *False positive*

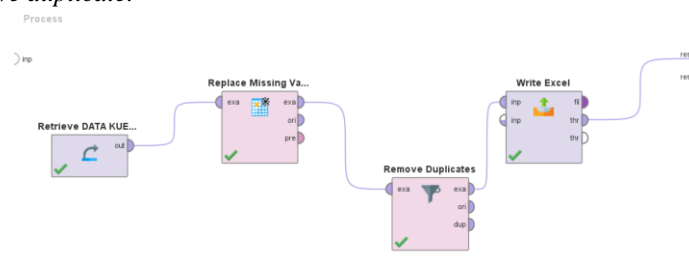
FN : *False negative*

### III. Hasil dan Pembahasan

#### A. Hasil Penelitian

##### 1. *Cleaning Data*

Tahap pertama melakukan pengumpulan menggunakan Tools Rapidminer yang dari data kuesioner para responden sebanyak 200 data yang kemudian dilakukan *replace missing* dengan menghilangkan data yang kosong dan melakukan *remove duplicate* dengan menghilangkan data yang duplikat atau data yang sama sehingga diperoleh data yang akan digunakan sebanyak 193 setelah melewati proses *replace missing* dan *remove duplicate*.



Gambar 2. Tampilan Proses *Cleaning Data*

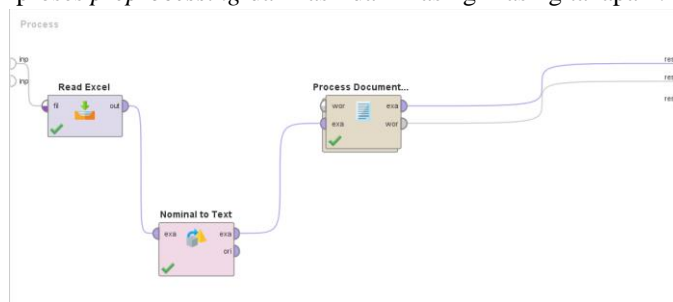
Row No.	Label	Apa alasan anda
1	Negatif	Karna Corona sudah tidak ada
2	Positif	Karena sebagai persyaratan perjalanan udara maupun darat saya setuju
3	Positif	Ingin mencoba
4	Positif	setuju vaksin mendukung program pemerintah
5	Negatif	Tidak perlu
6	Positif	Melengkapi proses vaksin sebelumnya
7	Positif	Karena untuk kesehatan saya setuju
8	Negatif	Corona sudah tdk ada
9	Positif	Ikut saja karna gratis
10	Negatif	Saya takut. Vaksin ketiga saja sudah cukup
11	Negatif	Sudah tidak banyak kasus covid saya pikir tidak perlu lagi dan saya lihat sudah tidak diwajibkan memakai masker ditempat umum
12	Negatif	Saya takut sudah berapa kali vaksin
13	Negatif	Tidak pernah ikut vaksinasi dari pertama Sampai sekarang

Gambar 3. Tampilan Data *Cleaning*

Gambar 3 merupakan tampilan hasil *cleaning* data yang menunjukkan data hasil *cleaning* setelah melewati proses *replace missing values* dan *remove duplicate* yang terdiri dari 193 data.

2. *Pre-Processing*

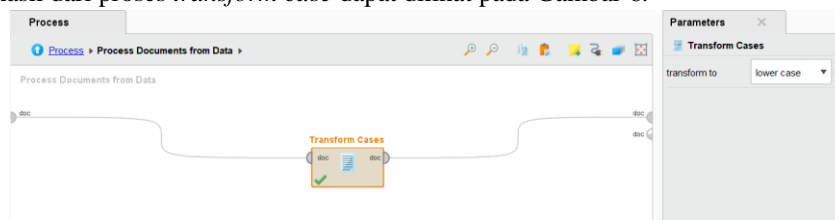
Setelah melewati tahap *cleaning* selanjutnya dilakukan proses *preprocessing*. Proses *preprocessing* dalam *Rapid Miner* menggunakan operasi *Process Document from Data* dan memiliki sub proses lainnya diantaranya *Transform Cases*, *Tokenize*, *Filter Token (by length)*, *Stemming*, dan *Filter Stopwords (dictionary)* sebelum masuk ke operator *Process Document Data* dimana terlebih dahulu melalui proses *Nominal to Text* yang bertujuan untuk memfokuskan operasi proses *document from data* pada atribut teks. Berikut desain proses *preprocessing* dan hasil dari masing-masing tahapan :



Gambar 4. Tampilan Desain *Preprocessing*

a. *Transform Cases*

Pada tahap ini dilakukan proses *Transform Cases* dengan mengubah teks menjadi huruf kecil dengan menggunakan operasi *Transform Cases* dengan *transform to: lower case* seperti pada Gambar 5 kemudian hasil dari proses *transform case* dapat dilihat pada Gambar 6.



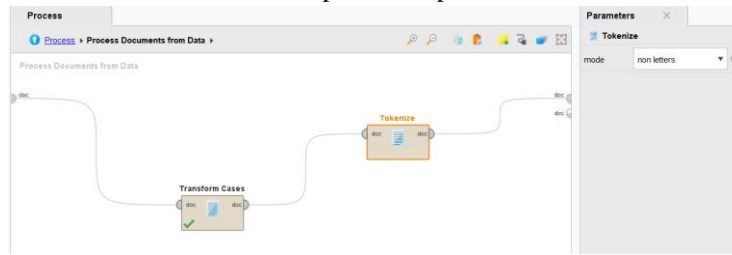
Gambar 5. Tampilan Desain Operasi *Transform Cases*

Word	Attribute Name	Total Occur...	Docum...	Negatif	Positif
agar dapat membantu me...	agar dapat membantu memustikan t...	1	1	0	1
agar dapat meningkatkan ...	agar dapat meningkatkan kekebalan ...	1	1	0	1
agar imunitas tubuh kita m...	agar imunitas tubuh kita meningkat d...	1	1	0	1
agar imunitas tubuh kita m...	agar imunitas tubuh kita meningkat d...	1	1	0	1
agar lebih aman	agar lebih aman	1	1	0	1
agar mendapat izin ke luar ...	agar mendapat izin ke luar kota atau ...	1	1	0	1
agar meningkatkan imunita...	agar meningkatkan imunitas tubuh d...	1	1	0	1
agar tubuh kuat melawan ...	agar tubuh kuat melawan virus coron...	1	1	0	1
agar tubuh lebih kebal dari ...	agar tubuh lebih kebal dari virus	1	1	0	1
aku udah vaksin 2	aku udah vaksin 2	1	1	1	0
alasan nya adalah untuk m...	alasan nya adalah untuk meningkatka...	1	1	0	1

Gambar 6. Tampilan Hasil Proses *Transform Case*

b. *Tokenize*

Pada tahap ini memecah sekumpulan kalimat menjadi kata, sekaligus menghilangkan karakter atau simbol dan tanda baca dengan menggunakan operator *tokenize* dengan *mode non letters* seperti pada Gambar 7 dan kemudian hasil dari *tokenize* dapat dilihat pada Gambar 8.



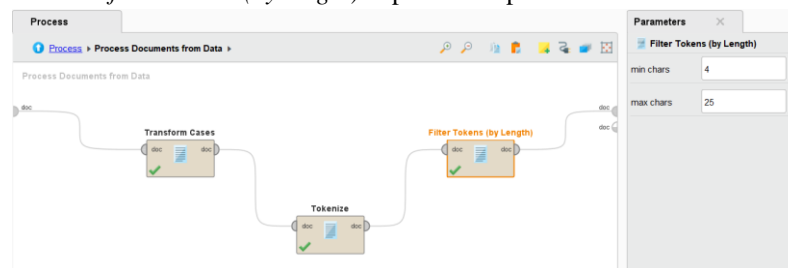
Gambar 7. Tampilan Desain Operasi *Tokenize*

Word	Attribute N...	Total Occure...	Documen...	Negatif	Positif
ada	ada	17	17	14	3
adalah	adalah	2	2	0	2
administrasi	administrasi	1	1	1	0
agar	agar	11	11	0	11
aja	aja	1	1	0	1
akan	akan	4	4	4	0
akses	akses	1	1	0	1
aku	aku	1	1	1	0
alasan nya	alasan nya	1	1	0	1
allah	allah	1	1	0	1

Gambar 8. Tampilan Hasil Proses *Tokenize*

c. *Filter tokens (by length)*

Pada tahap ini pengambilan kalimat yang penting dengan jumlah kata yang diinginkan dengan menyaring kata yang dibutuhkan dan menghilangkan kata yang tidak dibutuhkan. Pada proses ini dilakukan proses *filter tokens (by length)* dengan minimal chars: 4 dan maksimal chars: 25 yang bertujuan membatasi minimal dan maksimal karakter atau huruf dalam satu kata seperti pada Gambar 9. Kemudian hasil dari *filter tokens (by length)* dapat dilihat pada Gambar 10.



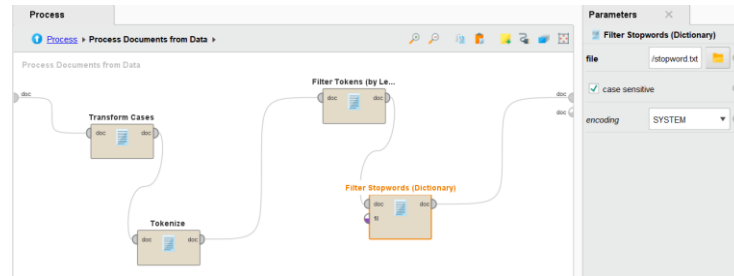
Gambar 9. Tampilan Desain Proses *Filter Tokens (By Length)*

Word	Attribute Name	Total Occurences	Document Occurences	Negatif	Positif
adalah	adalah	2	2	0	2
administrasi	administrasi	1	1	1	0
agar	agar	11	11	0	11
akan	akan	4	4	4	0
akses	akses	1	1	0	1
alasan nya	alasan nya	1	1	0	1
allah	allah	1	1	0	1
aman	aman	1	1	0	1
anjuran	anjuran	1	1	0	1
antibody	antibody	2	2	0	2
apabila	apabila	1	1	0	1

Gambar 10. Tampilan Hasil Proses *Filter Tokens (By Length)*

d. *Filter Stopword*

Pada tahap ini melakukan pengambilan kata dari hasil token yang membuang kata yang tidak penting, seperti kata “ke”, “di”, “dari” dihilangkan. Pada proses ini melakukan proses *filter stopwords* dengan mempersiapkan terlebih dahulu file kata yang tidak memiliki arti yang nantinya akan diinputkan ke dalam proses filter stopwords. Berikut ini adalah desain dan hasil dari proses filter stopwords yang ditunjukkan pada Gambar 11 dan Gambar 12.



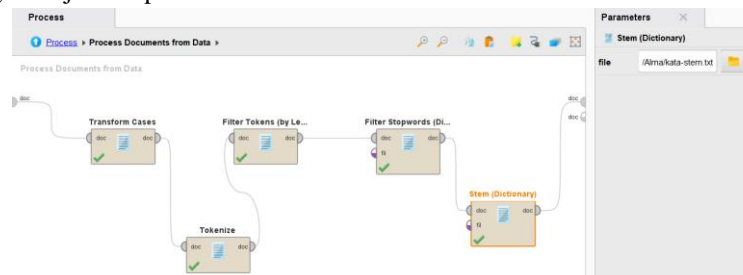
Gambar 11. Tampilan Desain Proses *Filter Stopwords*

Word	Attribute Name	Total Occurences	Document Occurences	Negatif	Positif
administrasi	administrasi	1	1	1	0
akses	akses	1	1	0	1
alasannya	alasannya	1	1	0	1
allah	allah	1	1	0	1
aman	aman	1	1	0	1
anjuran	anjuran	1	1	0	1
antibody	antibody	2	2	0	2
arahan	arahan	1	1	0	1

Gambar 12. Tampilan Hasil Proses *Filter Stopword*

e. *Stemming*

Pada tahap ini mengambil kalimat dalam bentuk kata dasarnya saja dan kata imbuhan dari tiap kalimat dihilangkan, seperti kata “persyaratan” menjadi kata “syarat”. Pada tahap ini menggunakan operator *Stem (Dictionary)* dengan mempersiapkan terlebih dahulu file kata dasar yang memiliki imbuhan yang nantinya akan diinputkan ke dalam proses stemming. Berikut ini adalah desain dan hasil dari proses stemming yang ditunjukkan pada Gambar 13 dan Gambar 14.



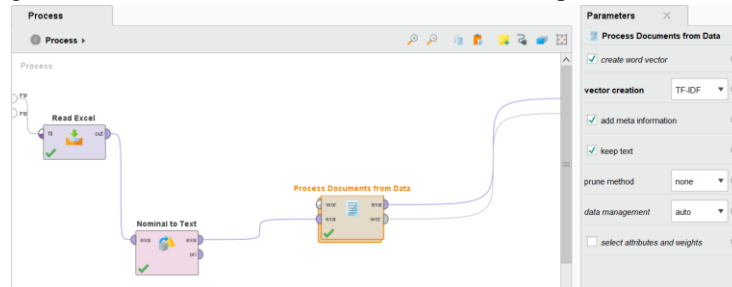
Gambar 13. Tampilan Desain Proses *Stemming*

Word	Attribute Name	Total Occurences	Document Occurences	Negatif	Positif
administrasi	administrasi	1	1	1	0
akhir	akhir	1	1	1	0
akses	akses	1	1	0	1
alasan	alasan	1	1	0	1
allah	allah	1	1	0	1
aman	aman	3	3	1	2
anjur	anjur	2	2	0	2
antibody	antibody	2	2	0	2
arah	arah	1	1	0	1
atas	atas	1	1	0	1
atur	atur	4	4	0	4

Gambar 14. Tampilan Hasil Proses *Stemming*

### 3. Pembobotan TF-IDF

Proses pembobotan TF-IDF ini dimulai dengan menghitung tiap *term* yang ada pada setiap dokumen. Kemudian proses dilanjutkan dengan menghitung jumlah dokumen yang memiliki term tertentu (DF). Setelah itu proses menghitung *Inverse Document Frequency* (IDF) dan yang terakhir nilai TF dikalikan dengan nilai IDF. Berikut adalah desain dari hasil dari proses TF-IDF sebagai berikut :



Gambar 15. Tampilan Desain Proses TF-IDF

Contoh perhitungan TF-IDF adalah sebagai berikut:

$$TF: \frac{\text{jumlah banyak kata dalam suatu dokumen}}{\text{jumlah kata pada dokumen}}$$

Contoh :

Term = cara urus administrasi untuk divaksin

$$TF_{(\text{administrasi, D148})} = \frac{1}{5} = 0.2$$

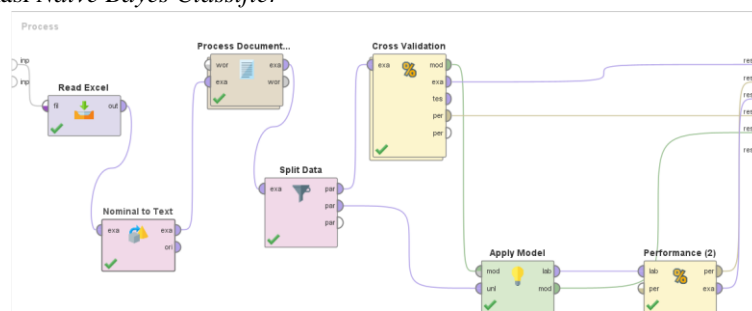
$$idf_t = \log\left(\frac{1+n}{1+df_t}\right) + 1 = \log\left(\frac{1+93}{1+1}\right) + 1 = 2.97$$

$$w_{t,d} = w_{tf} \times idf_t = 0.2 \times 2.96 = 0.592$$

Row No.	Label	text	administrasi	akhir	akses	alasan	allah	aman	anjur	antibody	arah
134	Positif	divaksin lanjut booster pertama kedua rasa perlin	0	0	0	0	0	0	0	0	0
135	Positif	sehat	0	0	0	0	0	0	0	0	0
136	Negatif	takut tubuh	0	0	0	0	0	0	0	0	0
137	Positif	untuk jaga kebal tubuh virus	0	0	0	0	0	0	0	0	0
138	Negatif	merasa skeptis divaksin beri pemerintah	0	0	0	0	0	0	0	0	0
139	Negatif	merasa takut efek sampingnya divaksin	0	0	0	0	0	0	0	0	0
140	Negatif	merasa janggal laksana divaksin beri	0	0	0	0	0	0	0	0	0
141	Negatif	orang larang untuk divaksin	0	0	0	0	0	0	0	0	0
142	Positif	merasa divaksin penting hindar serang sakit wab.	0	0	0	0	0	0	0	0	0
143	Negatif	mlk biaya untuk bayar divaksin	0	0	0	0	0	0	0	0	0
144	Positif	patuh percaya atur pemerintah untuk divaksin	0	0	0	0	0	0	0	0	0
145	Positif	orang tenaga sehat anjur untuk divaksin	0	0	0	0	0	0	0.580	0	0
146	Negatif	cara urus administrasi untuk divaksin	0.592	0	0	0	0	0	0	0	0
147	Positif	slap untuk divaksin	0	0	0	0	0	0	0	0	0
148	Negatif	merasa divaksin	0	0	0	0	0	0	0	0	0
149	Negatif	tunda untuk divaksin sibuk	0	0	0	0	0	0	0	0	0
150	Positif	dapat izin kota luar negeri dapat suntik divaksin	0	0	0	0	0	0	0	0	0

Gambar 16. Tampilan Hasil Proses TF-IDF

### 4. Proses Klasifikasi *Naïve Bayes Classifier*



Gambar 17. Tampilan Proses Klasifikasi

Gambar 17 merupakan tampilan data proses klasifikasi *naïve bayes* pada rapid miner, terlebih dahulu dengan memasukkan data penelitian yang berupa file *excel* yang telah melalui proses *cleaning* data berisi komentar yang mengandung sentimen negatif dan positif kemudian dihubungkan ke *operator nominal to text* agar data uji bisa terproses ke tahap *pre-processing* dan selanjutnya menyambungkan ke operator *split data* untuk membagi data training dan data testing dengan persentase 80% : 20%, kemudian dilakukan *cross validation* untuk memasukkan metode *naïve bayes* dan menghitung uji tingkat validitas data.



Row No.	Label	prediction	confidence	confidence	text	administrasi	akhir	akses	alasan	allah	aman
1	Negatif	Negatif	1.000	0.000		0	0	0	0	0	0
2	Negatif	Negatif	1	0	takut divaksin ketiga	0	0	0	0	0	0
3	Negatif	Positif	0	1	merasa sehat sbknm faksin per	0	0	0	0	0	0
4	Negatif	Negatif	1	0	laksana vaksin booster	0	0	0	0	0	0
5	Positif	Positif	0	1	program pemerintah	0	0	0	0	0	0
6	Negatif	Negatif	1	0	corona	0	0	0	0	0	0
7	Positif	Positif	0	1	kama vaksin gratis	0	0	0	0	0	0
8	Positif	Positif	0	1	bar sehat	0	0	0	0	0	0
9	Positif	Positif	0	1	bar sehat	0	0	0	0	0	0
10	Negatif	Negatif	1.000	0.000	susu	0	0	0	0	0	0
11	Positif	Positif	0	1	pikir khusus untuk usa tahun a	0	0	0	0	0	0
12	Positif	Positif	0	1	tambah kebal tubuh	0	0	0	0	0	0
13	Positif	Positif	0	1	untuk tambah tingkat kebal tub.	0	0	0	0	0	0
14	Negatif	Negatif	1	0	cemas efek samping garis tren.	0	0	0	0	0	0
15	Positif	Negatif	1.000	0.000	covid	0	0	0	0	0	0
16	Positif	Positif	0	1	wabp pemerintah	0	0	0	0	0	0

Gambar 18. Hasil Prediksi *Naive Bayes*

Gambar 18 merupakan hasil prediksi dari metode *naive bayes* yang mengklasifikasikan data uji ke dalam kelompok kelas positif dan kelas negatif berdasarkan nilai *confidence* negatif dan *confidence* positif tiap kalimat. Adapun contoh perhitungan dari hasil klasifikasi data uji adalah sebagai berikut:

$$P(\text{label}=\text{positif}) = \frac{\text{jumlah}_{\text{label}=\text{positif}}}{(\text{jumlah}_{\text{label}=\text{positif}} + \text{jumlah}_{\text{label}=\text{negatif}})}$$

$$P(\text{label}=\text{positif}) = \frac{102}{102+90} = \frac{102}{192} = 0.532$$

$$P(\text{label}=\text{negatif}) = \frac{\text{jumlah}_{\text{label}=\text{negatif}}}{(\text{jumlah}_{\text{label}=\text{positif}} + \text{jumlah}_{\text{label}=\text{negatif}})}$$

$$P(\text{label}=\text{negatif}) = \frac{90}{102+90} = \frac{90}{192} = 0.468$$

Dari perhitungan tersebut dapat ditentukan bahwa nilai probabilitas untuk label positif adalah 0.532 dan label negatif adalah 0.468

Selanjutnya menghitung *Conditional Probabilities* yaitu peluang masing-masing nilai terhadap label positif dan negatif.

Contoh kalimat: **takut divaksin ketiga**

$$P(\text{kata}=\text{takut} | \text{label}=\text{positif}) = \frac{\text{jumlah}_{\text{kata}=\text{takut} | \text{label}=\text{positif}}}{\text{jumlah}_{\text{kata}=\text{takut}}} = \frac{0}{13} = 0$$

$$P(\text{kata}=\text{takut} | \text{label}=\text{negatif}) = \frac{\text{jumlah}_{\text{kata}=\text{takut} | \text{label}=\text{negatif}}}{\text{jumlah}_{\text{kata}=\text{takut}}} = \frac{13}{13} = 1$$

$$P(\text{kata}=\text{divaksin} | \text{label}=\text{positif}) = \frac{\text{jumlah}_{\text{kata}=\text{divaksin} | \text{label}=\text{positif}}}{\text{jumlah}_{\text{kata}=\text{divaksin}}} = \frac{24}{66} = 0.363$$

$$P(\text{kata}=\text{divaksin} | \text{label}=\text{negatif}) = \frac{\text{jumlah}_{\text{kata}=\text{divaksin} | \text{label}=\text{negatif}}}{\text{jumlah}_{\text{kata}=\text{divaksin}}} = \frac{42}{66} = 0.636$$

$$P(\text{kata}=\text{ketiga} | \text{label}=\text{positif}) = \frac{\text{jumlah}_{\text{kata}=\text{ketiga} | \text{label}=\text{positif}}}{\text{jumlah}_{\text{kata}=\text{ketiga}}} = \frac{1}{3} = 0.33$$

$$P(\text{kata}=\text{ketiga} | \text{label}=\text{negatif}) = \frac{\text{jumlah}_{\text{kata}=\text{ketiga} | \text{label}=\text{negatif}}}{\text{jumlah}_{\text{kata}=\text{ketiga}}} = \frac{2}{3} = 0.66$$

Berdasarkan perhitungan yang telah dilakukan, hasil nilai probabilitas tersebut akan dihitung menggunakan teorema Bayes.

$$P(\text{positif}) = P(\text{label}=\text{positif}) \times P(\text{kata}=\text{takut} | \text{label}=\text{positif}) \times P(\text{kata}=\text{vaksin} | \text{label}=\text{positif}) \times P(\text{kata}=\text{ketiga} | \text{label}=\text{positif}) = 0.532 \times 1 \times 1 \times 0.363 = 0.1931$$

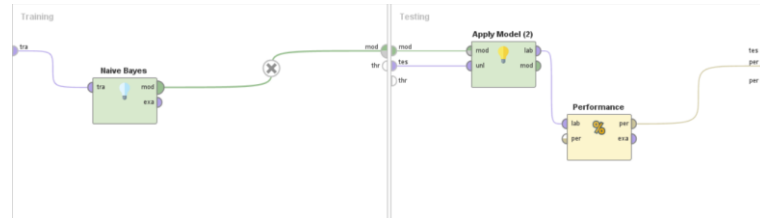
$$P(\text{negatif}) = P(\text{label}=\text{negatif}) \times P(\text{kata}=\text{takut} | \text{label}=\text{negatif}) \times P(\text{kata}=\text{divaksin} | \text{label}=\text{negatif}) \times P(\text{kata}=\text{ketiga} | \text{label}=\text{negatif})$$

$$= 0.468 \times 0 \times 0 \times 0.636 = 0$$

$$\text{Nilai probabilitas positif} = \frac{P(\text{positif})}{P(\text{positif}) + P(\text{negatif})} = \frac{0.1931}{0.1931+0} = 1$$

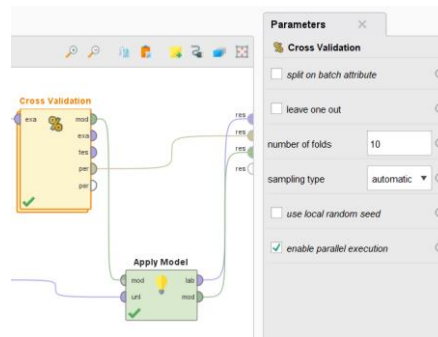
$$\text{Nilai probabilitas negatif} = \frac{P(\text{negatif})}{P(\text{positif}) + P(\text{negatif})} = \frac{0}{0+0.1931} = 0$$

##### 5. Perhitungan Klasifikasi Naïve Bayes



Gambar 19. Tampilan Proses Metode *Naive Bayes Classifier*

Gambar 19 merupakan tampilan proses dari metode *naïve bayes* dimana pada bagian training memasukkan operator algoritma metode *naïve bayes* sedangkan bagian testing memasukkan operator *apply model* dan *performance*.



Gambar 20. Pengujian *Cross Validation*

Gambar 20 merupakan tampilan pengujian *cross validation* dengan menggunakan k-fold=10 untuk melihat hasil *performance* dari nilai *accuracy*, *recall*, *precision*, dan *f1-score* berdasarkan hasil dari klasifikasi metode *naïve bayes*.

### PerformanceVector

```

PerformanceVector:
accuracy: 82.88% +/- 13.13% (micro average: 83.12%)
ConfusionMatrix:
True:  Negatif Positif
Negatif:    54     8
Positif:    18    74
precision: 81.51% +/- 12.82% (micro average: 80.43%) (positive class: Positif)
ConfusionMatrix:
True:  Negatif Positif
Negatif:    54     8
Positif:    18    74
recall: 90.14% +/- 11.46% (micro average: 90.24%) (positive class: Positif)
ConfusionMatrix:
True:  Negatif Positif
Negatif:    54     8
Positif:    18    74
AUC (optimistic): 0.944 +/- 0.063 (micro average: 0.944) (positive class: Positif)
AUC: 0.811 +/- 0.135 (micro average: 0.811) (positive class: Positif)
AUC (pessimistic): 0.800 +/- 0.141 (micro average: 0.800) (positive class: Positif)

```

Gambar 21. Tampilan Hasil *Performance Data Training*

Gambar 21 merupakan tampilan hasil *performance* dari data training, dimana pada bagian ini menampilkan hasil *accuracy* 83.12%, *presicion* 80.43% dan *recall* 90.42% dari data training dengan menggunakan fold=10.

## B. Pembahasan

### 1. Analisa Validitas Data

Pada tahap ini melakukan perhitungan untuk melihat nilai performance dari data testing yang telah diterapkan dimana hasil dari tahap ini diantaranya nilai akurasi, presisi, recall sebagai berikut:

#### a. Menghitung akurasi (*accuracy*)

$$\text{Akurasi} = \frac{\text{true positif} + \text{true negatif}}{\text{true positif} + \text{true negatif} + \text{false positif} + \text{false negatif}} =$$

$$= \frac{18+15}{18+15+3+2} = \frac{33}{38} = 86.84\%$$

#### b. Menghitung presisi (*precision*)

$$\text{Presisi} = \frac{\text{true positif}}{\text{false positif} + \text{true positif}} = \frac{18}{3+18} = \frac{18}{21} = 85.71\%$$

#### c. Menghitung recall (*recall*)

$$\text{Recall} = \frac{\text{true positif}}{\text{true positif} + \text{false negatif}} = \frac{18}{18+2} = \frac{18}{20} = 90.00\%$$

#### d. Menghitung F1-Score

$$\text{F1-Score} = 2 \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} = 2 \frac{90 \times 85.71}{90 + 85.71} = 2 \frac{7.713,9}{175,71} = 2 \times 43.90 = 87.80$$

### PerformanceVector

```

PerformanceVector:
accuracy: 86.84%
ConfusionMatrix:
True:  Negatif Positif
Negatif:    15     2
Positif:     3    18
precision: 85.71% (positive class: Positif)
ConfusionMatrix:
True:  Negatif Positif
Negatif:    15     2
Positif:     3    18
recall: 90.00% (positive class: Positif)
ConfusionMatrix:
True:  Negatif Positif
Negatif:    15     2
Positif:     3    18
AUC (optimistic): 0.953 (positive class: Positif)
AUC: 0.803 (positive class: Positif)
AUC (pessimistic): 0.786 (positive class: Positif)

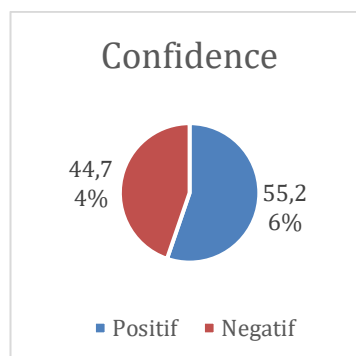
```

Gambar 22. Tampilan Hasil *Performance Vector Naive Bayes*

Gambar 22 merupakan tampilan hasil *performance vector naive bayes*, pada bagian ini menampilkan hasil performa dari *naive bayes* dengan berdasarkan nilai K-Fold = 10 yang terdiri *accuracy*, *precision* dan *recall* sehingga didapatkan *performance accuracy* sebesar 86.84%, *precision* sebesar 85.71% dan *recall* sebesar 90%.

### 2. Visualisasi Data

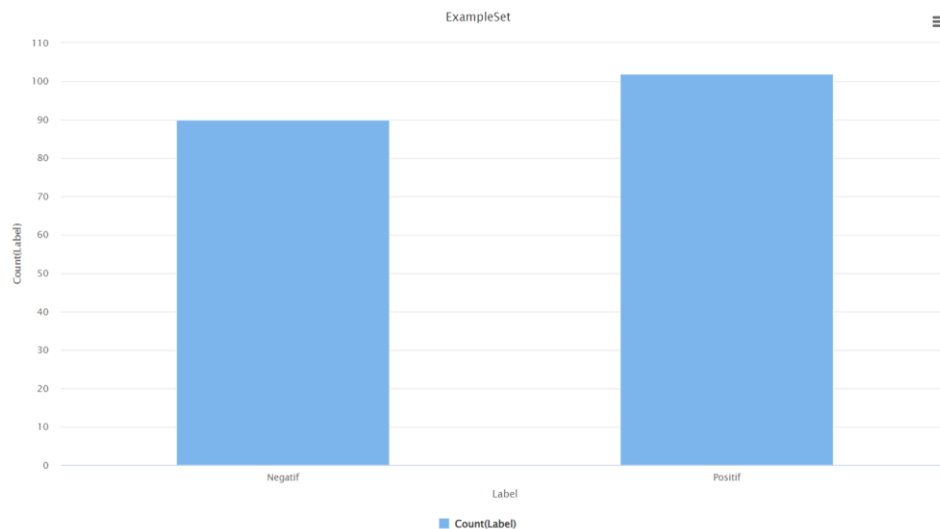
Pada tahap ini digunakan visualisasi diagram menggunakan diagram *pie* yang menunjukkan seberapa besar presentase sentimen positif dan sentimen negatif dari hasil analisis penelitian vaksinasi booster 2 menggunakan metode *Naive Bayes Classifier*.



Gambar 23. Diagram Hasil Prediksi Data Uji

Gambar 23 merupakan diagram hasil prediksi berdasarkan data uji yang menampilkan presentase yang menunjukkan bahwa komentar positif lebih dominan sebanyak 21 ( 55,26%) sentimen positif dan 17 (44,74%) sentimen negatif.

Name	Type	Missing	Statistics	Filter
Label Label	Binominal	0	Negative Negatif Positive Positif	Values Positif (102), Negatif (90)
Text text	Text	0	Least wajib un [...] ksin (1)	Most (5) Values (5), takut (4), ...[173 more]
administrasi	Real	0	Min 0	Max 0.592 Average 0.003
akhir	Real	0	Min 0	Max 0.913 Average 0.007
akses	Real	0	Min 0	Max 0.468 Average 0.002
alasan	Real	0	Min 0	Max 0.451 Average 0.002
allah	Real	0	Min 0	Max 0.535 Average 0.003
aman	Real	0	Min 0	Max 1 Average 0.013



Gambar 24. Diagram Hasil Pelabelan Sentimen

Gambar 24 merupakan diagram hasil dari pelabelan sentiment yang menunjukkan bahwa komentar positif lebih dominan dibandingkan dengan komentar negatif. Untuk komentar positif didapatkan sebanyak 102 sentimen (53.12%) sedangkan komentar negatif didapatkan sebanyak 90 sentimen (48.87%).

#### IV. Kesimpulan dan saran

Penelitian ini telah berhasil melakukan analisis sentimen menggunakan metode *Naïve Bayes Classifier* terhadap analisis sentimen masyarakat terhadap kebijakan pemerintah vaksinasi booster 2. Berdasarkan implementasi dari hasil evaluasi yang telah dilakukan. Maka dapat disimpulkan bahwa komentar positif lebih dominan dibandingkan dengan komentar negatif. Untuk komentar positif didapatkan sebanyak 21 ( 55,26%) sentimen, dan untuk komentar negatif didapatkan sebanyak 17 (44,74%) sentimen. Berdasarkan hasil pengujian performance data yang telah dilakukan menggunakan metode *Naïve Bayes Classifier* memperoleh *accuracy* sebesar 86.84, nilai *precision* sebesar 85.71% dan nilai *recall* sebesar 90%.

Adapun saran untuk penelitian selanjutnya agar lebih baik dan berkembang yaitu menggunakan jenis algoritma lebih dari satu kemudian dibandingkan jenis algoritma mana yang memiliki nilai akurasi yang tinggi. Data yang dikumpulkan bukan hanya melalui kuesioner saja, tetapi dikumpulkan dari berbagai macam media sosial yang memang pada konsidi sekaranag masyarakat menyuarakan isi hati dan opini mereka melalui era digital.

### Daftar Pustaka

- [1] A. Rozaqi, A. Triayudi, and R. T. Aldisa, "Analisis Sentimen Vaksinasi Booster Berdasarkan Twitter Menggunakan Algoritma Naïve Bayes dan K-NN," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 184, 2022, doi: 10.30865/json.v4i1.4907.
- [2] R. Pujiyanto, D. Yusup, and T. N. Padilah, "Analisis Sentimen Opini Publik Tentang Vaksin Booster Menggunakan Metode Support Vector Machine dan firefly Algorithm," *Jurnal Ilmiah Wahana Pendidikan, Desember*, vol. 2022, no. 23, pp. 363–373, 2022.
- [3] D. T. Anggraeni, "Analisis Sentimen Vaksinasi Booster Covid-19 pada Platform Twitter Menggunakan Metode Naïve Bayes," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 12, no. 2, p. 113, 2022, doi: 10.36448/expert.v12i2.2812.
- [4] R. T. Aldisa and P. Maulana, "Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
- [5] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [6] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [7] Y. Yunitasari and A. R. Putera, "Analisis Sentimen Masyarakat di Twitter Terkait Pandemi Covid-19," *Smatika Jurnal*, vol. 11, no. 01, pp. 22–26, 2021, doi: 10.32664/smatika.v11i01.520.
- [8] T. Com, "Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naïve Bayes Classifier," vol. 21, no. 3, pp. 644–654, 2022.
- [9] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Health Information Management Journal*, vol. 8, no. 2, pp. 100–109, 2020.
- [10] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding of Engineering*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [11] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 7, no. 1, p. 32, 2021, doi: 10.24014/coreit.v7i1.12945.
- [12] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [13] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [14] L. Syafie, F. Umar, A. Mude, H. Darwis, Herman, and Harlinda, "Missing Data Handling Using The Naive Bayes Logarithm (NBL) Formula," *IEEE Xplore*, pp. 1–4, 2018.
- [15] L. Ma and D. B. Ajipratama, "in Indonesia on Twitter Using the LSTM Algorithm," vol. 24, no. 2, pp. 161–172, 2022.