



## Analisis Sentimen Pada Media Sosial Twitter Mengenai Tanggapan Vaksinasi COVID-19 Menggunakan Metode *Naive Bayes*

Agus Suwarno<sup>1</sup>, Andriani<sup>2</sup>, Syaiful Rohman<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Industri Universitas Pelita Bangsa

Korespondensi email: [agussuwarno@pelitabangsa.ac.id](mailto:agussuwarno@pelitabangsa.ac.id)

### Abstraksi

*Coronavirus has become a pandemic in the world and has spread to almost all parts of the world, including in Indonesia. Many negative impacts resulted from the spread of covid-19 in Indonesia, so the government took action to implement vaccination programs to reduce the spread of covid-19. The response from the public to the vaccination program is quite diverse on social media Twitter, some agree and some disagree. this type of comment on twitter in the form of unstructured text in large numbers resulted in the missed information about COVID-19 vaccination that is useful from a set of text documents. It also knows the sentiment of twitter users manually can be detrimental to time and effort. This study aims to determine public sentiment towards vaccination. The data used are 1000 tweets, using two keywords, namely "vaccine + covid". The data is then divided into training data and test data. Classification was carried out using the Naïve Bayes method and using 10-Fold Cross Validation. The classification results from the Naïve Bayes method get an average accuracy of 73.75%.*

*Keywords: COVID-19, Twitter, Vaccination, Naive Bayes..*

### I. Pendahuluan

Ditemukannya vaksin COVID-19 menjadi babak baru penanganan pandemi. Harapan pandemi akan segera usai dengan terbentuknya kekebalan masyarakat secara menyeluruh, ada juga kerinduan akan normalnya kembali berbagai segi kehidupan. Vaksin COVID-19 kini telah tiba di Indonesia. Pengujian keamanan dan efektivitasnya telah dan sedang dilakukan pihak terkait. Penegasan Presiden mengenai vaksinasi kepada seluruh masyarakat dengan gratis. Bahkan, untuk

membangun kepercayaan publik, Presiden Jokowi memberikan contoh lebih dulu untuk divaksin. Media yang digunakan masyarakat untuk memberikan respon dan opininya salah satunya adalah media twitter, jenis komentar pada twitter berbentuk *unstructured text* dalam jumlah besar mengakibatkan terlewatnya informasi mengenai vaksinasi COVID-19 yang berguna dari sekumpulan dokumen teks tersebut. Menilai dan mengetahui sentimen secara manual dari pengguna twitter dapat merugikan waktu dan tenaga. Masih banyak masyarakat tidak

yakin dan tidak mau menerima vaksin COVID-19 dan rencana vaksinasi tersebut. Kemenkes pada September 2020 melakukan survei dan ditemukan ada 7,6% masyarakat menolak dan 27% ragu akan vaksinasi.

Penggunaan media yang semakin pesat dan diberbagai bidang dan berbagai macam lapisan masyarakat. Media tidak hanya sebagai media untuk bersosialisasi, dan berkomunikasi tetapi juga untuk menyampaikan aspirasi dan mempresentasikan tentang apa saja yang terjadi dan dirasakan oleh masyarakat. Penyampaian pendapat masyarakat yang melalui media sosial Twitter dalam menanggapi atau merespon adanya wabah virus Corona. Hal ini dapat sebagai acuan untuk mengetahui sentimen masyarakat terhadap dampak yang dirasakan oleh masyarakat Indonesia mengenai vaksinasi virus Corona.

Analisis sentimen bertujuan untuk dapat mengelompokkan polaritas dari teks, hal tersebut dilakukan untuk mengetahui bagaimana polaritas dari suatu tweet apakah opini bersifat positif, negatif atau netral. Dengan demikian dapat diketahui siapa pemberi opini yang paling banyak mendapatkan respon dari pengguna twitter. Analisis sentimen digunakan untuk mengungkapkan opini atau publik terhadap suatu isu, kepuasan pelayanan, kebijakan, *cyber bullying*, memprediksi harga saham dan analisis pesaing berdasarkan data tekstual yang ada.

## II. Metode Penelitian

Pada penelitian ini, objek penelitian yang digunakan adalah data

tweet berbahasa Indonesia yang terdapat pada media sosial twitter. Tweet yang digunakan ialah tweet-tweet yang mengandung sentimen positif dan sentimen negatif yang diperoleh dari tanggal 8 April 2021 - 10 April 2021. Dari masing-masing sentimen tersebut diambil 500 data per sentimen, sehingga total tweet yang digunakan sebagai data berjumlah 1000 tweet.

Pencarian data “vaksin + covid” dilakukan dengan menggunakan *keyword*, kemudian dilakukan pelabelan secara otomatis yaitu menggunakan kamus *lexicon* berbahasa Indonesia, tweet yang dipilih kemudian disimpan ke file *teks* dengan format *csv*. Kemudian file teks tersebut digunakan sebagai input untuk diolah lebih lanjut.

Berikut contoh tweet dengan sentimen positif:

Bersyukur, Alhamdulillah keluarga Persyarikatan Muhammadiyah sangat antusias untuk mendapatkan vaksinasi yang kedua. Dan semoga hal ini isa menjadi bagian dari ikhtiar kita yang sungguh-sungguh untuk bisa mengendalikan laju pandemi Covid-19 di negara Indonesia <a href="https://t.co/1pUVDzd1En">https://t.co/1pUVDzd1En</a> #Vaksin #COVID19
---

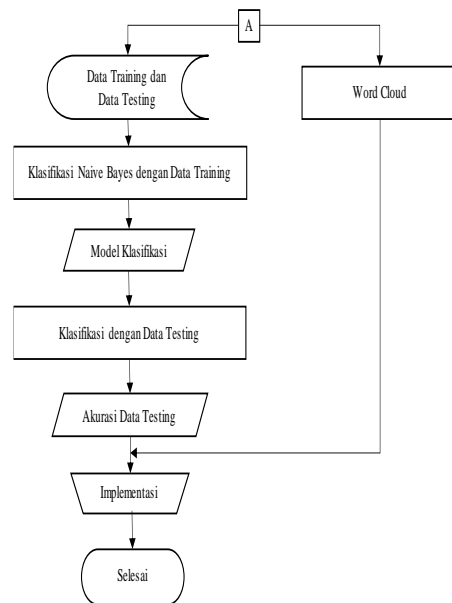
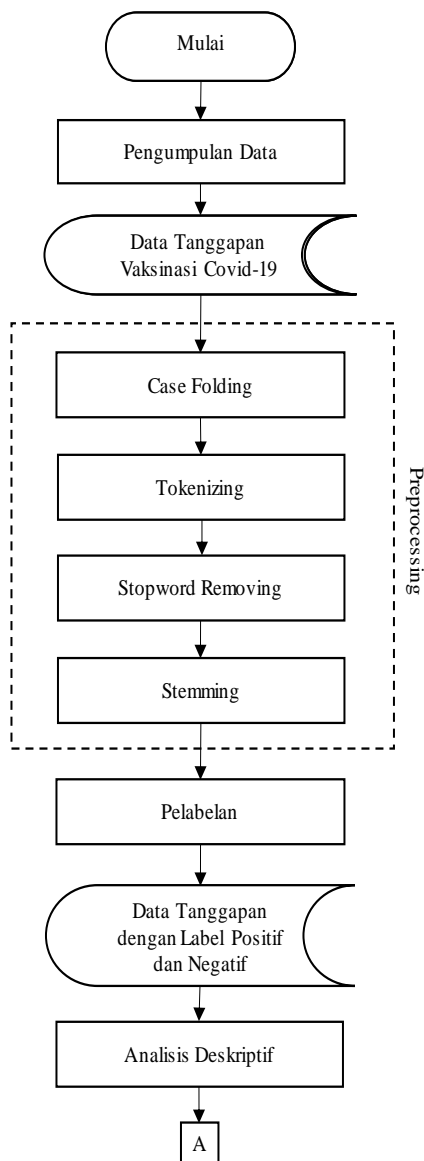
Berikut contoh tweet dengan sentimen negatif:

Ada beberapa orang yang mengeluhkan kakinya bengkak setelah di vaksin. Benarkah karena efek vaksin COVID-19? <a href="https://t.co/xGaLHsjzn">https://t.co/xGaLHsjzn</a>
---

Kerangka pemikiran adalah garis besar dari keseluruhan langkah-langkah dalam penelitian yang sedang dilakukan, kerangka pemikiran menjadi acuan untuk melakukan setiap tahapan

yang sedang dilakukan dalam penelitian. Berikut ini gambaran keseluruhan penelitian dalam bentuk diagram kerangka pemikiran:

Berikut ini *flowchart* yang menunjukkan tahapan penelitian mulai dari pengumpulan data hingga diperoleh kesimpulan.



**Gambar 1.** Tahapan Penelitian

Tahapan dimulai dengan mengumpulkan data, data dikumpulkan melalui media sosial twitter dengan menggunakan *keyword* “vaksin + covid”, kemudian data tersebut disimpan. Setelah itu, dilakukan *preprocessing* dimana tahapan-tahapan *preprocessing* tersebut meliputi *case folding*, *tokenizing*, *stopword removing*, dan *stemming*. Data yang telah melalui tahapan *preprocessing* akan diberi pelabelan data secara otomatis dengan menggunakan kamus *lexicon*. Pembagian pelabelan data dibagi dalam tiga kelas kategori yaitu tanggapan negatif, tanggapan positif, dan tanggapan netral. Data hasil pelabelan tersebut disimpan dan dilakukan analisis untuk mengetahui persentase tanggapan positif, tanggapan negatif dan tanggapan netral dari seluruh data yang diperoleh.

Kemudian data dibagi ke dalam data latih (*data training*) dan data uji (*data testing*), *data training* digunakan untuk

melatih *machine learning* agar dapat bekerja dengan baik dan *data testing* digunakan untuk menguji kinerja *machine learning*. Kemudian data yang telah melalui Tahapan *preprocessing* yang telah dilalui dan pelabelan juga divisualisasikan ke dalam sebuah *word cloud*.

Kemudian data dibagi menjadi dua bagian yaitu data latih (*data training*) dan data uji (*data testing*), serta dilakukan klasifikasi menggunakan algoritma *Naive Bayes*. Proses klasifikasi dimulai dengan menggunakan data training sehingga diperoleh model klasifikasi, selanjutnya model klasifikasi yang telah diperoleh digunakan untuk melakukan klasifikasi pada *data testing* sehingga diperoleh nilai akurasi dari hasil klasifikasi.

Menggunakan data yang telah divisualisasikan ke dalam *word cloud* akan menghasilkan kata dominan yang paling sering dibahas oleh pengguna Twitter. Setelah semua tahapan telah terlewati, maka dilakukan interpretasi terhadap hasil yang diperoleh.

Adapun langkah-langkah dalam tahapan penelitian yang dilakukan dalam proses penelitian ini:

#### a. Pengumpulan Data

Pada tahap pengumpulan data ini akan dijelaskan tentang dari mana data penelitian dihasilkan.

#### b. *Preprocessing*

Pada tahap ini terdiri dari proses *Case Foolding*, *Tokenizing*, *Stopword Removing*, dan *Stemming*.

#### c. Pembagian Data

Tahapan ini menentukan jumlah *data training* dan *data testing* yang akan digunakan

#### d. Simulasi perhitungan

Tahap ini menjelaskan tentang perhitungan.

Metode *scrapping* dilakukan untuk mengambil data menggunakan Twitter API dari tanggal 8 April 2021 - 10 April 2021 dengan kata kunci vaksin + covid". Twitter API merupakan aplikasi yang diciptakan oleh pihak twitter dengan tujuan agar mempermudah pihak developer untuk mengakses informasi web twitter. Pendaftaran API digunakan untuk mengonfirmasi kepada pihak twitter agar memberikan izin menjelajahi lebih luas terkait dengan data yang berkaitan dengan twitter.

Pada penelitian ini, menggunakan metode dengan teknik *web scrapping* dan alat yang digunakan untuk *crawling* data adalah R/R Studio. *Crawling data* pada twitter diperlukan sebuah kode yang didapat dari twitter API untuk memperoleh akses data twitter tersebut.

Kemudian melakukan registrasi dan bergabung dengan twitter API sehingga didapatkan beberapa kode berupa *consumer key*, *consumer secret*, *access token* dan *access key* dari twitter. Kode API tersebut digunakan sebagai jembatan antara twitter dengan aplikasi lainnya yang digunakan untuk proses integrasi antara twitter API dengan R.

Kemudian data yang diperoleh sudah dalam bentuk *csv* kemudian dilakukan tahap *preprocessing*, tahap ini dilakukan bertujuan untuk

membersihkan data-data dari *noise* dan pembenahan bahasa seperti menghilangkan singkatan, bahasa gaul, serta menghapus kata yang tidak perlu, data awal yang didapatkan berupa data yang tidak terstruktur maka perlu dilakukan tahap *preprocessing* supaya data tersebut dapat dianalisis.

Hasil dari *crawling* data mentah atau data yang diperoleh masih terdapat unsur simbol, URL dan sebagainya yang tidak mempunyai arti pada kalimat tersebut. Hal ini dapat menyulitkan para pembaca untuk menemukan topik atau pembahasan informasi terkait. Dari hal tersebut diperlukan proses *cleaning* guna membersihkan data sehingga pembaca dapat mengetahui informasi dengan mudah. Proses *cleaning data* adalah proses merapihkan dan membersihkan kalimat dari kata-kata yang tidak mempunyai arti sehingga lebih mudah dan cepat dalam mendapatkan informasi yang didapat. Dengan proses *cleaning* mendapatkan hasil pada tabel berikut:

**Tabel 2.** Hasil Proses *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Berarti tidak salah kalau kita sebut virus covid-19 tidak terlalu berbahaya. Tp ya jaga2 saja supaya tdk terinfeksi. Better selalu patuh saja pada prokes. Buang jaub2 mindset bahwa vaksin (China) adalah solusi. @KetumProDEMnew @kandargalang <a href="https://t.co/KkVsSkHZ8u">https://t.co/KkVsSkHZ8u</a>	Berarti tidak salah kalau kita sebut virus covid tidak terlalu berbahaya Tp ya jaga saja supaya tdk terinfeksi Better selalu patuh saja pada prokes Buang jaub mindset bahwa vaksin China adalah solusi KetumProDEMnew kandargalang

a. *Case Folding*

*Case folding* adalah tahap proses pengubahan huruf kapital menjadi

huruf non kapital atau semuanya menjadi huruf kecil. Pada Tabel dibawah ini terdapat huruf kapital, dimana huruf tersebut yang dirubah pada proses *case folding*.

b. *Tokenizing*

Tahapan *tokenizing* merupakan proses memisahkan kata di dalam dokumen menjadi potongan kata yang tidak saling berpengaruh dan disebut *token* untuk kemudian dapat diidentifikasi. Pada tabel berikut merupakan contoh dari proses *tokenizing*.

c. *Stopword Removing*

*Stopword Removing* merupakan tahapan untuk pengambilan kata-kata yang penting. Proses *Stopword Removing* menggunakan fungsi *stopword* (menghapus kata tidak penting). Contoh *stopword* yaitu “yang”, “dan”, “ke”, “dari”, “di” dan lainnya. Kata-kata tersebut sangat berfrekuensi tinggi dan dapat ditemukan pada hampir setiap kalimat. *Stopword* /menghapus kata bertujuan mengurangi ukuran indeks dan waktu proses serta dapat mengurangi *noise*.

d. *Stemming*

*Stemming* merupakan perubahan berbagai kata imbuhan menjadi sebuah kata dasar, baik imbuhan deduktif maupun imbuhan induktif. Pada tabel berikut merupakan contoh dari proses *Stemming*.

Data set dalam penelitian ini adalah tanggapan mengenai vaksinasi COVID-19 pada media sosial twitter dari tanggal 8 April 2021 - 10 April 2021. Data

tanggapan yang diperoleh sebanyak 1833 tanggapan. Penentuan jumlah data training dan testing dengan perbandingan rasio untuk masing-masing data positif dan negatif sebanyak 500 data tanggapan positif dan 500 data tanggapan negatif, sehingga total data tanggapan yang digunakan sebesar 1000 data tanggapan sedangkan untuk data netral tidak digunakan karena tidak memberikan informasi yang penting. Pada penelitian ini untuk seluruh data dibagi ke dalam 5 kombinasi perbandingan rasio *data training* dan *data testing*

Data set dalam penelitian ini adalah tanggapan mengenai vaksinasi COVID-19 pada media sosial twitter dari tanggal 8 April 2021 - 10 April 2021. Data tanggapan yang diperoleh sebanyak 1833 tanggapan. Penentuan jumlah *data training* dan *testing* dengan perbandingan rasio untuk masing-masing data positif dan negatif sebanyak 500 data tanggapan positif dan 500 data tanggapan negatif, sehingga total data tanggapan yang digunakan sebesar 1000 data tanggapan sedangkan untuk data netral tidak digunakan karena tidak memberikan informasi yang penting. Pada penelitian ini untuk seluruh data dibagi ke dalam 5 kombinasi perbandingan rasio *data training* dan *data testing* seperti pada tabel dibawah ini.

**Tabel 3.** Perbandingan Rasio

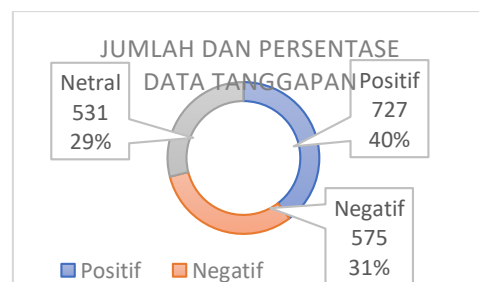
No	Jumlah Data	Rasio Training: Testing (%)	Data Training	Data Testing
1	1000	20:80	200	800
2		40:60	400	600
3		60:40	600	400
4		80:20	800	200
5		90:10	900	100

Simulasi perhitungan yang dilakukan adalah :

- a. Simulasi TF-IDF
- b. Simulasi *Naïve Bayes*
- c. Simulasi Perhitungan *Kinerja Machine Learning*

### III. Hasil dan Pembahasan

Analisis data tanggapan digunakan untuk mengetahui gambaran umum mengenai data tanggapan tentang vaksinasi COVID-19 setelah melalui *text preprocessing* dan pelabelan secara otomatis oleh kamus *lexicon*. Data tanggapan yang diperoleh pada 8 April 2021 - 10 April 2021 ini dikategorikan ke dalam tiga kelas kategori yaitu tanggapan positif, tanggapan negatif dan tanggapan netral. Secara keseluruhan, gambaran mengenai tanggapan yang diperoleh berdasarkan masing-masing kategori ditunjukkan oleh grafik berikut ini.



**Gambar 2.** Tanggapan

Klasifikasi ini dilakukan dengan menggunakan algoritma *Naive Bayes*. Klasifikasi dilakukan untuk mengetahui hasil kinerja *machine learning* dalam melakukan klasifikasi terhadap data tanggapan. Total data tanggapan yang telah diperoleh dibagi ke dalam kombinasi data dengan menggunakan perbandingan rasio, untuk selanjutnya



dilakukan klasifikasi dengan algoritma yang telah ditentukan. Hasil klasifikasi tersebut disajikan dalam bentuk tabel.

Data training merupakan data yang digunakan pada fase training, sedangkan data testing merupakan data yang digunakan pada fase uji. *Data training* maupun *data testing* terbagi ke dalam dua kategori yaitu kategori tanggapan positif dan tanggapan negatif. Total keseluruhan data sebesar 1000 tanggapan, jumlah masing-masing data sebesar 500 tanggapan positif dan 500 tanggapan negatif. Pembagian jumlah data tanggapan diseimbangkan supaya hasil dari kinerja *machine learning* tidak terpacu pada mayoritas data yang jumlahnya lebih besar.

Pengujian kinerja *machine learning* difokuskan pada perbandingan rasio yang memiliki kinerja terbaik didasarkan pada tingkat akurasi yang tinggi. Perbandingan akurasi berdasarkan jumlah data uji pada algoritma *Naive Bayes* pada masing-masing rasio berdasarkan hasil evaluasi menggunakan *10-fold cross validation* ditunjukkan oleh tabel dibawah ini.

Dilakukan terhadap masing-masing klasifikasi kelas sentimen, adapun tujuan visualisasi adalah untuk mengekstraksi informasi berupa topik yang paling sering di bicarakan / diulas oleh pengguna Twitter, sehingga dari sekian banyak tanggapan yang ada, dapat diambil informasi yang dianggap penting berdasarkan kata yang paling sering muncul. Berikut penjelasan hasil visualisasi kata dari setiap klasifikasi kelas sentimen.



Gambar 3. Ulasan Positif



Gambar 4. Ulasan Negatif

#### IV. Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada pembahasan sebelumnya, penulis dapat menarik beberapa kesimpulan sebagai berikut:

1. Hasil analisis komentar Twitter mengenai vaksinasi COVID-19 dengan menggunakan *text mining* diperoleh data sejumlah 1833 tweet, didapat hasil klasifikasi positif sebesar 727 data, klasifikasi negatif sebesar 575 data dan klasifikasi netral sejumlah 531 data.
2. Hasil klasifikasi sentimen secara otomatis pada Twitter menggunakan *text mining* dengan menggunakan algoritma *Naive Bayes* diperoleh hasil rerata nilai akurasi sebesar 65,58% tanpa *cross validation* dan rerata akurasi sebesar 73,75% dengan menggunakan *10-cross validation*. Sedangkan untuk

visualisasi kata tanggapan positif mengenai vaksinasi COVID-19 sering dikaitkan dengan kata “terima”, “negara”, “papua”, “masyarakat”, “sehat”, “maju”, “dosis”, “otsus”, “suntik” dan “takut” sedangkan untuk tanggapan negatif sering dikaitkan dengan kata “pandemi”, “buka”, “laporcovid”, “sekolah”, “warga”, “koalisi”, “tolak”, “agustus”, “peta” dan “orang.

### Daftar Pustaka

- [1] M. Tyas, sheli janu setyaning Febianah, “TEMATIK - Jurnal Teknologi Informasi Dan Komunikasi Vol. 3, No. 1 Juni 2016,” Anal. Perbandingan Algoritm. Naive Bayes dan C.45 Dalam Klasifikasi Data Min. Untuk Memprediksi Kelulusan, vol. 3, no. 1, pp. 86–99, 2021.
- [2] D. R. Beniac, A. Andonov, E. Grudeski, and T. F. Booth, “Architecture of the SARS coronavirus prefusion spike,” *Nat. Struct. Mol. Biol.*, vol. 13, no. 8, pp. 751–752, 2006, doi: 10.1038/nsmb1123.
- [3] S. Syamaidzar, “Review Vaksin Covid-19,” *Res. Gate*, no. July, pp. 1–15, 2020.
- [4] KemenkesRI, “Vaksinasi COVID-19 Lindungi Diri, Lindungi Negeri,” *Kementeri. Kesehat. RI*, vol. 9, pp. 22–50, 2021.
- [5] B. M. Pintoko and K. M. L., “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier,” *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [6] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, “Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.