

# ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP PROGRAM VAKSIN COVID-19

ANALYSIS OF YOUTUBE USER SENTIMENT ON COVID-19 VACCINE PROGRAM

**Ferian Fauzi Abdulloh\*<sup>1</sup>, Iqbal Rilo Pambudi<sup>2</sup>**

Universitas Amikom Yogyakarta ; Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman,  
Daerah Istimewa Yogyakarta 55281

Jurusan Informatika, Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, Yogyakarta  
e-mail: \*<sup>1</sup>ferian@amikom.ac.id, <sup>2</sup>[iqbal.pambudi@students.amikom.ac.id](mailto:iqbal.pambudi@students.amikom.ac.id)

## **Abstrak**

Vaksinasi telah mulai dilakukan pemerintah Indonesia per tanggal 13 Januari 2021 menjadi momen yang penting. Vaksinasi menjadi pilihan pemerintah Indonesia untuk menekan grafik penularan virus Covid-19. Indonesia telah menyiapkan dosis vaksin sebesar 371 juta vaksin corona. Namun, isu tentang vaksin juga beredar luas di masyarakat. Rumor tentang vaksin yang belum aman dan kurangnya sosialisasi menimbulkan pro dan kontra terhadap vaksinasi ini. Maka dari itu mencari tahu opini masyarakat tentang vaksinasi ini menjadi opsi untuk menentukan sentimen masyarakat terhadap vaksin. Media sosial menjadi salah satu sumber opini masyarakat dalam menuangkan pendapatnya. Pengolahan opini tersebut dapat menjadi sebuah alternatif untuk menentukan respon publik terhadap suatu peristiwa tertentu. Menurut HootSuite Indonesia tahun 2021 Youtube menjadi social media dengan pengguna terbanyak. Dengan pertimbangan tersebut, Youtube menjadi sebuah pilihan untuk menjadi sumber dataset. Selanjutnya dataset itu diolah dan diproses dengan metode Support Vector Machine data yang diolah tersebut akan dijadikan model klasifikasi untuk melakukan pengklasifikasian teks dari komentar Youtube dengan topik bahasan “vaksin covid-19”. Hasil dari pengklasifikasian tersebut diharapkan dapat menjadi informasi maupun masukan kepada pihak tertentu untuk dijadikan pertimbangan.

**Kata kunci**— Vaksin, Youtube, Support Vector Machine

## **Abstract**

The vaccination carried out by the Indonesian government on January 13, 2021 is an important moment. Vaccination is the choice of the Indonesian government to suppress the graph of Covid Virus Infections and transmissions. Indonesia has prepared a vaccine dose of 371 million corona vaccines. However, the issue of vaccines is also widely circulated in the community. Rumors about unsafe vaccines and lack of socialization lead to pros and cons of this vaccination. Therefore, finding out public opinion about vaccination is an option to determine public sentiment towards vaccines. Social media is a source of public opinion in expressing their opinions. Opinion processing can be an alternative to determine the public's response to a particular event. According to HootSuite Indonesia in 2021 Youtube will become the social media with the most users. With these considerations, Youtube is an option to become a dataset source. Furthermore, the dataset will be processed and processed. Support Vector Machine method will be used as a classification model to classify the text from Youtube comments with the topic "covid-19 vaccine". The results of the classification are expected to provide information and input to certain parties for consideration

**Keywords**—Vaccine, Youtube, Support Vector Machine

## 1. PENDAHULUAN

Media sosial adalah sebuah fitur pada internet yang memungkinkan pengguna internet saling berbagi informasi. Media sosial juga merupakan salah satu alat dimana seseorang dapat saling

terhubung dengan segala hal. Dengan media sosial pengguna dapat saling berinteraksi dengan orang lain di seluruh dunia dengan berbagi foto, video pesan singkat dan sebagainya. Media sosial memberikan dampak yang signifikan terhadap keberlangsungan hidup para generasi millennial hal ini terjadi karena masyarakat khususnya anak muda telah tumbuh dan terintegrasi dengan jaringan media sosial. Hasil riset menunjukkan 53 responden yang terdiri dari 35 responden perempuan dan 18 laki-laki menunjukkan mereka telah menggunakan sosial media mulai dari umur 15-20 tahun . Pertumbuhan media sosial juga dibarengi dengan perkembangan smartphone yang semakin mendorong pertumbuhan pengguna media social[1].

Sekarang ini terdapat banyak sekali sosial media yang ada di Internet. Facebook , Twitter , Instagram, Youtube dan tentunya beberapa sosial media yang lainnya. Dilansir dari HootSuite Indonesia tahun 2021, Youtube menempati posisi pertama sebagai sosial media dengan pengguna terbanyak yaitu sebesar 107 juta pengguna. Berdasarkan statistik tersebut Youtube dapat menjadi sumber data yang berasal dari komentar para netizen. Kolom komentar menjadi wadah untuk 1 memberikan opini, respon dan tanggapan terhadap suatu konten video yang diunggah di Youtube. Dengan mengukur seberapa banyak komentar negatif dan positif pada suatu konten di sosial media dapat menjadi sebuah alternatif untuk menilai sentimen publik terhadap konten tersebut, salah satunya vaksin covid-19.

Seperti yang diketahui, vaksin adalah *hot topic* yang sering diperdebatkan akhir-akhir ini. Banyak stigma yang muncul terhadap vaksin di masyarakat. Maka dari itu untuk mengetahui polaritas masyarakat terhadap vaksin ini perlu dilakukannya analisis sentimen dengan mengolah data komentar. Namun dengan banyaknya jumlah data komentar dan tidak memungkinkannya untuk menghitung satu persatu maka digunakanlah pendekatan Text Mining untuk mengolah banyaknya komentar tersebut apakah dikategorikan sebagai komentar positif atau negatif memakai metode Support Vector Machine dataset komentar tersebut akan diolah sehingga menghasilkan pola klasifikasi yang nantinya akan digunakan untuk mengukur tingkat polaritas terhadap topik vaksin covid-19 pada platform Youtube.

## 2. METODE PENELITIAN

### 2.1 Tahapan Review

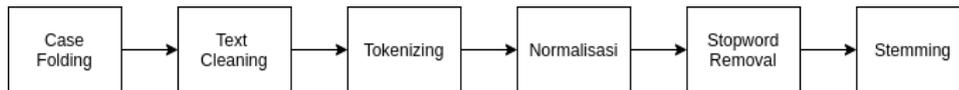
Analisis sentimen atau biasa kita kenal dengan opinion mining ialah sebuah bidang studi yang mempelajari dan menganalisis sebuah pendapat, sikap dan emosi seseorang terhadap suatu peristiwa atau entitas tertentu. Ada dua metode dasar dalam analisis sentimen, yang pertama adalah tingkat dokumen dan yang kedua adalah tingkat kalimat. Tingkatan dokumen menganalisa sentimen pada keseluruhan isi dokumen dan tingkatan kalimat hanya menganalisa sentimen pada tingkat kalimat[2]. Analisis sentimen biasa digunakan untuk mengukur tingkat kecenderungan seseorang terhadap suatu hal dalam sebuah pembicaraan. Penggalan informasi dalam opini dapat menjadi sebuah studi untuk mencari kecenderungan seseorang dan dapat dimanfaatkan dalam beberapa bidang. Dalam bidang bisnis, analisis sentimen dapat dimanfaatkan untuk mengukur opini atau pendapat konsumen terhadap produk atau layanan yang mereka miliki, sehingga dapat dilakukan evaluasi produk ataukah layanan yang fit dengan kebutuhan konsumen.

Text mining sendiri lebih kepada proses penggalian informasi dengan tingkat kualitas dan akurasi yang tinggi dari teks. Informasi yang berkualitas ini diperoleh melalui penemuan pola dan tren (seperti mempelajari pola statistik)[3]. Tujuan dari text mining itu sendiri adalah untuk menemukan beberapa kata yang menjadi substansi utama suatu dokumen sehingga dapat menganalisis hubungan antar dokumen[4]. Text Mining sendiri adalah bagian dari data mining yang akan memproses banyak data atau teks dalam jumlah yang besar. Proses Text mining sendiri tentunya menjadi sama dengan data mining, kecuali memang pada beberapa metode tertentu dan jenis data

yang dianalisis, seperti halnya data teks yang tidak terstruktur, terstruktur sebagian maupun yang telah terstruktur, maupun komentar-komentar dari banyak media[5].

### 2.1.1 Preprocessing

Preprocessing meliputi tahapan case folding, text cleaning, tokenizing, normalisasi, stopword removal dan stemming. Case folding akan mengubah huruf kapital menjadi huruf kecil. Text cleaning akan menghilangkan tanda baca, karakter, huruf tunggal yang dianggap tidak memiliki arti dalam pemrosesan kata sehingga menghasilkan kalimat yang bersih. Kemudian akan dilakukan normalisasi yang diperlukan untuk mengubah kata-kata singkatan atau kata yang tidak umum lainnya sehingga menjadi seragam.



Gambar 1. Alur Preprocessing

### 2.1.2 Pembobotan Term

TF-IDF ialah pembobotan yang biasanya dipergunakan dan seringkali dipakai dalam penentuan bobot dari suatu data dalam text mining. Bobot sendiri adalah ukuran secara statistik yang digunakan untuk menakar seberapa penting kata tertentu dalam sebuah dokumen dalam kumpulan corpus[6]. Metode ini adalah cara yang efisien dan sederhana yang masih banyak digunakan sampai sekarang untuk menimbang bobot sebuah term[7].

### 2.1.3 Support Vector Machine

Support Vector Machine (SVM) pada awalnya diperkenalkan oleh Guyon, Boser dan Vapnik pada 1992 di Annual Workshop on Computational Learning Theory. SVM sendiri ialah sebuah metode dalam pembelajaran supervised learning yang berdasar pada prinsip Structural Risk Minimization (SRM) yang digunakan untuk mencari hyperplane yang terbaik yang memisahkan 2 kelas[2]. SVM akan berusaha mencari hyperplane terbaik untuk memisahkan dua buah kelas. Adapun Hyperplane pemisah terbaik diantara 2 kelas dapat dihitung dengan cara mengukur margin hyperplane kemudian juga dicari titik maksimalnya. Hyperplane yang baik merupakan hyperplane yang mempunyai jarak terbesar ke titik data training terdekat dari setiap kelas, karena biasanya semakin besar margin, maka semakin rendah pula error generalisasi dari pemisah. Dan adapun margin yaitu jarak dari suatu titik vektor di suatu kelas terhadap hyperplane[8]. Mencari hyperplane yang optimal merupakan tujuan dari SVM itu sendiri dengan cara memaksimalkan margin dengan nilai jarak antara hyperplane dan titik terdekatnya dengan:

$$\min \frac{1}{2} \|w\|^2$$

Dengan subjek

$$y_i(w \cdot x_i + b) \geq 1$$

dimana  $x_i$  adalah data input dan  $y_i$  adalah nilai keluaran dari data  $x_i$ ,  $w$  dan  $b$  adalah parameter yang akan dicari nilainya.

### 2.1.4 Dataset

Sumber data dari penelitian ini berasal dari komentar video youtube yang diambil menggunakan fitur Youtube Data API v3 yang disediakan oleh Youtube dan dapat digunakan secara umum. Dataset ini kemudian akan dilakukan pelabelan secara manual untuk memisahkan sentimen yang kemudian akan dipelajari oleh SVM.

Tabel 1. Pelabelan Dataset

<b>Komentar</b>	<b>Label</b>	<b>Sentimen</b>
awas jangan mau itu pembodohan	0	Negatif
Gw kagak mau, takut nyawa gw dikurangin sama si vaksin, mending dikerokin aja dah	0	Negatif
Yap bisnis besar sebentar lg akan dimulai	1	Netral
Apakah vaksin sudah bisa d temukan,, semoga indonesia bisa normal kembali seperti biasa	1	Netral
Semoga secepatnya ada vaksin, kakekku mati karna corona	2	Positif
Aamiin semoga Pandemi Virus Corona ini cepat berakhir dan berlalu tahun 2021 normal kembali	2	Positif

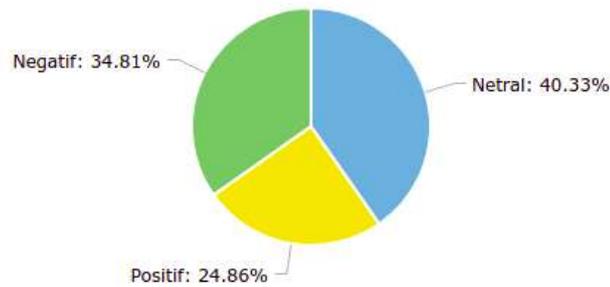
### 3. HASIL DAN PEMBAHASAN

Pada tahapan ini data latih yang akan digunakan adalah sebesar 1448 data dengan rincian 504 dataset negatif, 584 dataset netral, 360 dataset positif.

Tabel 2. Sampel data latih

<b>Komentar</b>	<b>Label</b>
Jangan Ngaku Cebong Kalau Belum Berani Daftar Jadi Relawan Yang Di Vaksin Covid-19 Buatan China	Negatif
"Wow buang duwit aja... Semoga dasa2ku di ampuni ,, aamiin"	Negatif
Yeaah ladang bisnis lagi sm negeri tirai bambu.	Negatif
yang bikin virus cina yang buat obatnya harus beli dr cina juga	Negatif
Hebat bikin virus corona mereka ciptakan vaksin nya.. Politik ekonomi	Negatif
Giliran beritanya kaya gini yg asal muasalny bikin Ekonomi jadi DOWN , malah pada Sepi	Netral
Ha....Ha.....Ha..... gue sampai ketawa lihat pada komentarnya seru banget.....	Netral
Kenapa ga ada perusahaan yg mau meneliti nasi kucing dan jamu yg terbukti miliki berjuta manfaat	Netral
kami butuh uang,,makan,, tuk lanjutkan hidup	Netral
KETAHUILAH , TIDAK SEMUA TUBUH MANUSIA KUAT TERHADAP VAKSIN .....	Netral
"semoga Benar, Akurat dan Tepat untuk Covid-19 virus ."	Positif

Yang mau vaksin ya vaksin, yang gak mau ya terserahhh.. duh pada ribut aja	Positif
saya jg mau pak di vaksin tp sy di hongkong	Positif
Vaksin virus Corona harus ditemukan supaya ekonomi lancar lagi	Positif
Semoga vaksin nya bisa membuat warga indonesia semuanya sehat dan bisa normala lagi seperti dulu	Positif



Gambar 2. Grafik Dataset

Kemudian dataset tersebut akan dilakukan sampling dengan persebaran data 1:1:1 sehingga setiap kelas akan berjumlah 360 dataset. Kemudian dataset yang telah di sampling akan memasuki tahapan preprocessing untuk dilakukan pembersihan kata. Proses pembersihan kata ini akan mengakibatkan beberapa dokumen akan kosong nilainya dikarenakan kata yang terkandung pada dokumen tersebut dianggap tidak berarti dan dihilangkan pada tahapan preprocessing. Untuk menghindari kekosongan dokumen ini diperlukan lagi penyaringan dataset untuk membuang dataset yang tidak memiliki nilai atau kosong.

Dataset yang telah melewati tahapan preprocessing menghasilkan term yang siap untuk diolah dan dilakukan pembobotan. Untuk mengetahui persebaran kata dan kata apa saja yang sering muncul dapat menggunakan Wordcloud



Gambar 3. Wordcloud Sentimen

Dari WordCloud diatas dapat diketahui kata apa saja yang banyak bermunculan pada masing-masing kelas sentimen. Tahapan selanjutnya adalah pembobotan term dengan TF-IDF. Pembobotan ini menggunakan library Scikit-Learn pada bahasa python. Sebelum dilakukannya TF-IDF dataset akan dibagi antara data latih dan data uji dengan perbandingan 70:30.

Tahapan selanjutnya adalah pengklasifikasian dengan Support Vector Machine. Pengujian dilakukan dengan 4 kernel yaitu kernel linear, poly, RBF, dan sigmoid dengan parameter  $C=2$ ,  $\gamma = 2$ ,  $\text{degree}=3$  dan  $\text{decision\_function\_shape} = \text{'ovo'}$  (One versus One). Penggunaan one versus one dipilih dikarenakan data yang dilatih merupakan multiclass. Dari pelatihan data tersebut didapatkan hasil sebagai berikut dengan pembagian data latih dan uji sebesar 70% : 30%. Pengujian ini dilakukan dengan 2 tahapan, yang pertama menguji dengan 3 kelas sentimen berupa positif, netral dan negatif. Kemudian pengujian yang kedua dilakukan dengan 2 kelas sentimen berupa positif dan negatif.

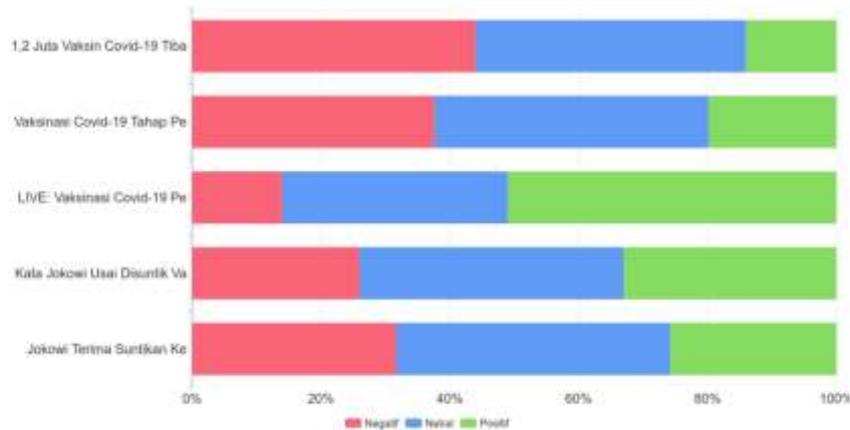
Tabel 3. Perbandingan Algoritma A dan Algoritma B

Kernel	Akurasi								Rata-rata
	Uji 1	Uji 2	Uji 3	Uji 4	Uji 5	Uji 6	Uji 7	Uji 8	
linear	77%	75%	68%	76%	82%	80%	78%	70%	75.7%
rbf	77%	76%	71%	78%	80%	80%	77%	70%	76.12%
sigmoid	76%	71%	67%	70%	76%	75%	76%	69%	77.75
poly	75%	75%	71%	74%	81%	78%	77%	71%	75.75

Model yang telah dilatih akan diujikan dengan dataset yang baru. Dataset yang baru berasal dari komentar pada beberapa video yang berhubungan dengan topik vaksin covid19.

Tabel 4. Objek Pengujian

Judul Video	Channel Youtube	Jumlah Tayangan	Tanggal Upload	Jumlah Komentar	Like/Dislike
1,2 Juta Vaksin Covid-19 Tiba di*****	CNN Indonesia	341.577x	7 Desember 2021	718	2,29rb/443
Vaksinasi Covid-19 Tahap Pe****	CNN Indonesia	92,278x	2 Januari 2021	222	719/130
LIVE: Vaksinasi Covid-19 Perdana di Indo*****	Sekretariat Presiden	851,253x	13 Januari 2021	527	13rb/825
KataJokowi Usai Disuntik *****	KOMPAS TV	506,907x	13 Januari 2021	901	5,9rb/577
Jokowi Terima Suntikan *****	CNN Indonesia	70.708x	27 Januari 2021	266	1,7rb/119



Gambar 4. Hasil Pengujian Sentimen

#### 4. KESIMPULAN

Berdasarkan penelitian yang sudah dijabarkan pada bab sebelumnya, didapatkan kesimpulan Support Vector Machine terbukti baik digunakan untuk mengklasifikasikan teks terutama dalam mengklasifikasikan komentar positif dan negatif pada platform Youtube dengan topik “vaksin covid-19”. Hasil pengujian sentimen publik terhadap vaksin sebelum peristiwa vaksinasi terhadap Jokowi menunjukkan netizen memiliki sentimen negatif terhadap vaksin sebesar 41% dan sentimen positif 17%. Namun, peristiwa vaksinasi terhadap Jokowi dan setelahnya menunjukkan sentimen negatif terhadap vaksin menjadi 24% dan sentimen positif sebesar 36%. Dari 5 video yang diujikan hanya 1 video yang jumlah like nya sebanding, sedangkan 4 video yang lain berbanding terbalik.

#### 5. SARAN

Dataset yang digunakan perlu ditingkatkan dan perlu dilakukan pelabelan yang lebih akurat sehingga menghasilkan akurasi dan ketepatan data yang baik.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu penulis dalam menyelesaikan publikasi ini dalam bentuk apapun.

#### DAFTAR PUSTAKA

- [1] M. A. D. Y. S. Dhifa Nabila, Octavia Elvaretta, Ghoniyatu Zahira, 2020, *Peradaban Media Sosial di Era Industri 4.0*. Inteligencia Media (Intrans Publishing Group), Malang.
- [2] A. Rahmawati, A. Marjuni, and J. Zeniarja, 2017, “Analisis Sentimen Publik Pada Media Sosial Twitter Terhadap Pelaksanaan Pilkada Serentak Menggunakan Algoritma Support Vector Machine,” *CCIT J.*, vol. 10, no. 2, pp. 197–206, doi: 10.33050/ccit.v10i2.539.
- [3] N. Wayan and S. Saraswati, 2013, “Naïve bayes classifier dan support vector machines

- untuk sentiment analysis,” pp. 2–4,.
- [4] A. S. Aribowo, 2015, “Arsitektur Aplikasi Twitter Opinion Mining Untuk Mengetahui Sentimen Publik Terhadap Merek,” vol. 2015, no. November, pp. 14–20.
- [5] A. T. J. H, 2015, “Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining,” *Inform. UPGRIS*, vol. 1, pp. 1–9.
- [6] D. Maulina and R. Sagara, 2018, “Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency-Inverse Document Frequency,” *J. Mantik Penusa*, vol. 2, no. 1, pp. 35–40.
- [7] J. Ramos, 1975, “Using TF-IDF to Determine Word Relevance in Document Queries Juan,” *Urol. Clin. North Am.*, vol. 2, no. 1, pp. 29–48.
- [8] V. I. Santoso, G. Virginia, and Y. Lukito, 2017, “Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine,” *J. Transform.*, vol. 14, no. 2, p. 72, doi: 10.26623/transformatika.v14i2.439.