

ANALISIS SENTIMEN DAN TOPIK PERBINCANGAN NETIZEN INDONESIA MENGENAI KEBIJAKAN VAKSIN COVID-19

Akhmad Rizal Dzikrillah¹, Dyvia Oliviani²

^{1,2}Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Prof. DR. Hamka, Jl. Tanah
Merdeka No.6, Kampung Rambutan, Pasar Rebo, Jakarta Timur, 13830

¹ahmadrizaldzikrillah@gmail.com

²dyviaololiviani@gmail.com

Abstrak: Pemerintah Republik Indonesia dalam upaya menjalankan percepatan penanganan penyebaran Covid-19 di Indonesia memiliki rencana mengenai pelaksanaan vaksinasi covid-19 bagi Masyarakat Indonesia pada tahun 2020 akhir. Akan tetapi, masyarakat masih banyak yang merasa kebingungan terhadap keamanan, efek samping dan cara memperolehnya. Pengumpulan data respon dan opini masyarakat Indonesia terhadap kebijakan vaksin COVID-19 dilakukan dengan menggunakan teknik web scraping pada media sosial twitter dalam bentuk tweets. Kegiatan mengklasifikasikan tweets atau analisis sentimen pada penelitian ini dilakukan dengan metode lexicon-based atau berbasis kamus positif-negatif. Mengacu pada penelitian yang dilakukan sebelumnya, kegiatan mengelompokkan opini masyarakat di media sosial twitter dilakukan dengan menggunakan model Latent Dirichlet Allocation (LDA). Sentimen positif lebih besar dibandingkan dengan Sentimen negatif mengenai kebijakan vaksin Covid-19. Topik-topik yang mencuat dan populer pada perbincangan para pengguna social twitter adalah topik edukasi fungsi vaksin, harapan, serta kepercayaan masyarakat terhadap vaksinasi. Topik populer yang lain adalah pemberitahuan netizen mengenai kegiatan vaksinasi yang telah dia ikuti.

Kata Kunci: Analisis Sentimen, Vaksin Covid-19, Indonesia.

Abstract: The Government of the Republic of Indonesia in an effort to handle the handling of the spread of Covid-19 in Indonesia has a plan to implement a covid-19 vaccination for the Indonesian people in late 2020. However, many people are still confused about its safety, side effects and how to get it. The collection of response data and Indonesian public opinion on the COVID-19 vaccine policy was carried out using web scraping techniques on Twitter social media in the form of tweets. The activity of classifying tweets or sentiment analysis in this study was carried out using the lexicon-based method or positive-negative dictionary-based. Referring to the research conducted previously, the activity of grouping public opinion on social media twitter was carried out using the Latent Dirichlet Allocation (LDA) model. Positive sentiment is greater than negative sentiment regarding the Covid-19 vaccine policy. The topics that stick out and are popular in the conversations of social twitter users are the topic

of education on the function of vaccines, hopes, and public trust in vaccination. Another popular topic is netizen notifications regarding activities he has participated in.

Keywords: Sentiment Analysis, Covid-19 Vaccine, Indonesia.

I. PENDAHULUAN

Pemerintah Republik Indonesia dalam upaya menjalankan percepatan penanganan penyebaran Covid-19 di Indonesia memiliki rencana pelaksanaan vaksinasi covid-19 bagi Masyarakat Indonesia pada tahun 2021. Pemerintah menargetkan vaksinasi untuk 70% penduduk Indonesia dengan dua kali suntik dosis untuk mencapai *herd immunity* (kekebalan komunitas) oleh pemerintah[1].

Penerimaan masyarakat terhadap penggunaan vaksin merupakan faktor yang mendukung tercapainya target vaksinasi pemerintah. Menurut Davis [2], persepsi masyarakat terhadap suatu barang seperti vaksin mempengaruhi penerimaan masyarakat terhadap penggunaan barang tersebut. Oleh karena itu, untuk mengetahui tingkat penerimaan masyarakat terhadap vaksinasi, maka dibutuhkan riset guna menganalisis persepsi masyarakat Indonesia terhadap vaksin.

Salah satu cara untuk menganalisis persepsi masyarakat adalah dengan menggunakan media sosial. Dengan menggunakan teknologi komputer dan metode analisis sentimen [3], kita dapat mengumpulkan teks-teks opini netizen yang relevan dengan suatu topik, untuk kemudian memprosesnya, lalu menganalisis subjektifitas atau polaritas sentimen para netizen pada topik terkait.

Berdasarkan kebutuhan tersebut, maka peneliti melakukan penelitian untuk menganalisis sentimen dan topik perbincangan netizen Indonesia terkait vaksinasi covid 19 di media sosial.

Rumusan masalah dari penelitian ini adalah:

1. Bagaimanakah perbandingan antara sentimen positif dan negatif para pengguna media sosial twitter dalam merespon kebijakan vaksinasi Covid-19?
2. Opini dan topik yang populer diperbincangkan para pengguna media sosial twitter dalam merespon kebijakan vaksinasi Covid-19?

II. TEORI

A. Web Scrapping

Cara kerja *web scrapping* adalah mengambil data dari internet dan website, setelah itu menyimpan data-data tersebut dalam format *file spreadsheet* seperti Google Sheet atau Microsoft Excel. Dua cara umum yang dapat dilakukan

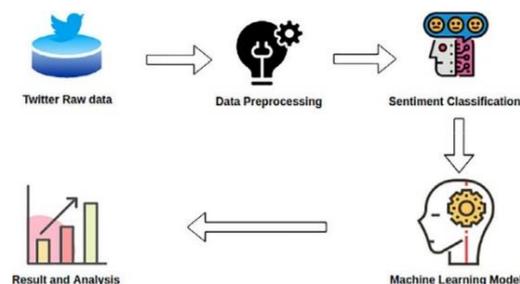
dalam proses *web scrapping*, yaitu secara manual dan otomatis. Cara *web scrapping* dengan metode otomatis dilakukan melalui bahasa pemrograman (koding), *browser extention*, atau aplikasi pembantu [4]. Salah satu teknik *web scrapping* secara otomatis adalah dengan menggunakan aplikasi berbasis *python* dan API *twitter*.

Referensi [5] menunjukkan bahwa API Twitter atau *Application Programming Interface (API)* twitter adalah suatu program atau aplikasi yang disediakan oleh twitter untuk mempermudah developer lain dalam mengakses informasi yang ada di website twitter. Setelah mendaftar developer akan mendapatkan *consumer key*, dan *access token* yang akan digunakan sebagai syarat otentifikasi dari aplikasi yang akan kita bangun. Tujuan dari otentifikasi adalah untuk hak akses developer dalam mengunduh data yang ada di twitter.

B. Analisis Sentimen

Sentiment analysis adalah cara untuk menarik subjektifitas dan polaritas emosi yang terkandung pada sebuah teks [3].

Langkah-langkah metode analisis sentimen dapat dilihat pada gambar 1.



Gambar 1. Metode analisis sentimen[6]

Langkah-langkah analisis sentimen berdasarkan referensi [6] adalah dengan cara mendapatkan sejumlah *tweet* yang relevan dengan topik yang dianalisis terlebih dahulu. Setelah itu tahap selanjutnya adalah melakukan pemrosesan data agar data dapat menjadi data yang mudah

dimengerti oleh mesin komputer. Setelah dilakukan pemrosesan data, maka dilakukanlah klasifikasi *tweet* berdasarkan sentimen. Setelah itu, kita menggunakan *machine learning* untuk menghasilkan hasil analisis sentimen dari sejumlah *tweet* yang kita dapatkan.

C. *Lexicon Based*

Lexicon dictionary [7] merupakan salah satu kamus yang digunakan dalam menganalisis data. *Dictionary* atau kamus *lexicon* digunakan *lexicon based* untuk menerapkan penilaian terhadap kata. Nilai polaritasnya disandingkan dengan kata-kata pada kamus. Yang perlu dilakukan sebelum melakukan analisis memakai *lexicon* yakni memilih kata yang akan dianalisis dari *corpus*. Kata tersebut dapat dipilih dengan melakukan *Part-Of-Speech Tagging* kemudian mencari kata-kata dengan tipe yang sudah ditentukan, misalnya *adjective* dan *adverb*.

Penentuan polaritas sentimen pada penelitian ini, menggunakan *matching* kata berdasarkan kamus *Lexicon-based approach*. Cara menentukan sentiment opini berdasarkan referensi adalah:

1. Memilih kata bersentimen : setiap kata di dalam sebuah kalimat akan diberi sebuah nilai yakni bernilai satu (1) jika kata bersentimen positif dan bernilai negatif satu (-1) jika kata bersentimen negatif.
2. Penerapan skor pada kalimat skor digunakan untuk memastikan apakah sebuah kalimat bersentimen positif atau bersentimen negatif. Skor kalimat didapat dari penjumlahan nilai dari kata bersentimen.

Nilai dari skor sentimen memastikan sentimen dari suatu kalimat dengan keadaan seperti berikut: Jika nilai sentimen > 0 , maka *tweets* bersentimen

positif. Jika nilai sentimen < 0 maka *tweets* bersentimen negatif; selain itu sentimen netral [8].

D. *Latent Dirichlet Allocation (LDA)*

Latent Dirichlet Allocation (LDA) [9] ialah suatu metode *topic modelling* yang digunakan untuk mengetahui topik-topik yang mencuat pada suatu dokumen. LDA adalah bentuk probabilistik generatif dari kumpulan tulisan yang disebut corpus. Dalam LDA, dokumen-dokumen merupakan objek pengamatan, sedangkan topik, penyebaran topik tiap dokumen, klasifikasi kata pada topik tiap dokumen adalah struktur tersembunyi. Ide dasar LDA adalah setiap dokumen direpresentasikan sebagai kombinasi acak dari topik yang terkandung, dan setiap topik mempunyai karakter yang direpresentasikan oleh penyebaran kata-kata yang ada di dalamnya.

Latent Dirichlet Allocation (LDA) adalah salah satu jenis dari *Unsupervised Machine Learning* [8] dan menggunakan prinsip *clustering* berbasis partisi [9]. Oleh karena itu, LDA membutuhkan pendefinisian jumlah kluster topik yang akan digunakan terlebih dahulu. Jumlah kluster yang telah ditentukan akan digunakan untuk melakukan pengelompokan topik tanpa memerlukan data *training*.

III. METODE PENELITIAN

Langkah-langkah dalam penelitian ini disusun berdasarkan metode analisis sentimen. Dengan menggunakan pendekatan *lexicon based* dan penggunaan LDA sebagai model untuk menganalisis sentimen. Serangkaian tahapan yang akan peneliti dapat lakukan pada Gambar 2.



Gambar 2. Metode penelitian

A. *Crawling Tweet*

Sebelum melakukan *crawling tweet*, peneliti membuat akun twitter terlebih dahulu. Dengan menggunakan akun twitter yang telah dibuat sebelumnya, peneliti membuat API twitter melalui twitter developer. API Twitter digunakan untuk melakukan *crawling* (penggalan) data Twitter setelah mendapatkan *key* dan *tokens*.

Peneliti melakukan *crawling* data menggunakan perangkat lunak berbasis *python* yang telah dipasangkan *key*, dan *tokens*. Untuk melakukan penggalan *tweet*, perangkat lunak diberi kata kunci “Efektifitas Vaksin Covid”, “Vaksin Covid-19” dan “Efek Vaksin Covid-19”. Waktu pengambilan *tweet* dilakukan dalam 2 waktu, yaitu pada Agustus 2021 mendapatkan 58 *tweets*, dan Desember 2021 mendapatkan 197 *tweets*.

B. *Data Preprocessing*

Data yang diperoleh dari proses *crawling* menggunakan *python* masih berupa data yang masih belum terstruktur dan dalam format *file json*. *File json* ditransformasi terlebih dahulu menjadi *file excel/csv*. Tujuan *file json* diubah

menjadi *file excel* ialah agar peneliti lebih mudah melakukan *preprocessing*.

Preprocessing diperlukan agar data dapat menjadi data yang mudah dimengerti oleh mesin komputer. Proses *preprocessing* dilakukan dengan bantuan *tools* Microsoft Excel. Alur *preprocessing* data twitter dapat dilihat pada gambar 3.



Gambar 3. Preprocessing

1) *Case Folding*: Semua huruf yang ada di dokumen menjadi huruf kecil dan yang diterima hanya huruf ‘a’ sampai ‘z’. Karakter selain huruf dihilangkan dan dianggap delimiter.

2) *Tokenizing*: *Tokenizing* adalah tahap pengurangan atau menghilangkan string input atas dasar tiap kata yang menyusunnya.

3) *Filtering*: *Filtering* merupakan hasil term yang didapatkan dari tahap pengambilan kata-kata. Dapat menerapkan algoritma *stoplist* (tidak menggunakan kata yang tidak terlalu penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* yaitu pendekatan *bag-of-word* dengan mengabaikan kata-kata yang tidak deskriptif.

4) *Stemming*: Pada tahap *stemming* peneliti mencari dasar kata dari setiap kata hasil *filtering*. Ditahapan ini proses pengambilan dilakukan dengan beragam bentuk kata ke dalam suatu representasi yang sama. Tahapan ini lebih mudah diterapkan pada teks dalam bahasa Inggris dan akan sulit digunakan pada teks berbahasa Indonesia.

5) *Translation*: Pada tahap *translation* peneliti menggunakan fitur yang terdapat pada Microsoft Office Word untuk menerjemahkan

istilah asing yang terdapat pada data. Data dalam bahasa asing diterjemahkan ke dalam bahasa Indonesia sehingga peneliti lebih mudah dalam menganalisis.

C. Klasifikasi Sentimen dengan Lexicon

Kegiatan mengklasifikasikan *tweets* atau analisis sentimen pada penelitian ini dilakukan dengan metode *lexicon-based* atau berbasis kamus positif-negatif. Kamus positif-negatif yang digunakan pada penelitian ini yaitu kamus positif-negatif yang dibuat oleh Liu, Hu & Cheng [7]. yang telah diterjemahkan ke dalam Bahasa Indonesia dan telah dilakukan penyesuaian Bahasa.

Dalam mengklasifikasikan *tweets* berbasis *lexicon*, langkah pertama yang dilakukan oleh peneliti adalah pemberian skor pada tiap kata. Setiap kata didalam sebuah kalimat akan diberi sebuah nilai yakni bernilai satu (1) jika kata bersentimen positif dan bernilai negatif satu (-1) jika kata bersentimen negatif.

Langkah kedua adalah memberikan skor pada masing-masing *tweet*. Skor kalimat didapat dari penjumlahan nilai dari kata bersentimen. Nilai dari skor sentimen memastikan sentimen dari suatu kalimat dengan keadaan seperti berikut: Jika nilai sentimen > 0 , maka *tweets* bersentimen positif. Jika nilai sentimen < 0 maka *tweets* bersentimen negatif; Jika nilai sentiment adalah 0, maka akan bernilai netral. Proses klasifikasi sentimen dilakukan dengan menggunakan bantuan Microsoft Excel.

D. Pemodelan Topik dengan LDA

Sebelum membuat pemodelan topik dengan LDA, peneliti terlebih dahulu membuat *Wordcloud* masing masing dari respon positif dan negatif data

agar lebih terlihat corpus yang ditampilkan dalam data penelitian ini. Pembuatan *wordcloud* menggunakan *software* berbasis python.

Setelah *corpus* ditampilkan dalam bentuk *Wordcloud*, peneliti melakukan pemodelan topik dengan menggunakan *Latent Dirichlet Allocation* (LDA). Proses analisis pegelompokan topik dengan LDA menggunakan bantuan *software* berbasis python. Karena LDA menggunakan prinsip *partition based clustering*, maka peneliti menentukan terlebih dahulu jumlah kluster dari *tweet-tweet* yang ada. Jumlah kluster topik pada penelitian ini adalah 12 kluster dengan tujuan agar model yang dibangun tidak menghasilkan topik yang saling tumpang tindih antar satu dengan yang lain, atau tidak terlalu mendetail sehingga banyak topik yang harus diinterpretasikan.

E. Analisis Sentimen dan Topik Perbincangan

Peneliti melakukan analisis perbandingan antara sentimen positif dan negatif. Peneliti juga menganalisis topik-topik pembicaraan yang mencuat di *wordcloud* masing masing dari sentimen positif dan negatif. Peneliti juga menganalisis hasil dari LDA yaitu 3 topik pembicaraan teratas yang paling sering dibicarakan netizen melalui kata-kata yang mewakili masing-masing topik.

IV. HASIL DAN PEMBAHASAN

Berdasarkan hasil klasifikasi sentimen berbasis *lexicon*, maka didapatkan perbandingan prosentase *tweet* yang cenderung bersentimen positif, netral, dan negatif. Perbandingan prosentase sentimen dapat dilihat pada Tabel 1.

TABEL 1. PERBANDINGAN PROSENTASE SENTIMEN

Sentimen	Jumlah Tweets	Prosentase
----------	---------------	------------

Dari Tabel 1, kita dapatkan bahwa topik yang paling populer dibicarakan di masyarakat adalah kluster opini yang diwakili oleh kata 'udah', 'booster', 'warganya', 'berharap', 'now', 'bertahan', 'semangat', 'kekebalan', 'percaya', 'persentase'. Topik kedua yang populer adalah kluster opini yang diwakili oleh kata-kata 'pemerintah', 'memperluas', 'provokator', 'guys', 'efektifitas', 'rantai', 'karna', 'dan' anak', 'mari'. Topik ketiga yang populer adalah kluster opini yang diwakili oleh kata-kata 'emang', 'vaksin', 'covid', 'dosis', 'dosis', 'tingkat', 'efektifitas', 'nya', 'aja', 'si', 'vaksin', 'covid', 'dosis', 'netral', 'nan', 'positif'.

Kata-kata representasi 3 topik peringkat teratas menurut analisis penulis berkaitan dengan harapan-harapan dan kepercayaan masyarakat menanggapi kebijakan vaksinasi pemerintah. Netizen juga membicarakan kegiatan vaksinasi pada dosis tertentu yang telah diikuti oleh sesama netizen dan informasi yang berkaitan dengannya.

V. KESIMPULAN

Berdasarkan analisis sentimen netizen terhadap kebijakan vaksinasi, maka dapat disimpulkan bahwa:

1. Opini netizen yang bersentimen positif terhadap kebijakan vaksinasi Covid-19 lebih besar daripada opini netizen yang bersentimen negatif dengan perbandingan hingga 13:1.
2. Salah satu topik yang mencuat dari percakapan netizen terkait vaksinasi adalah topik edukasi terkait fungsi vaksin kepada netizen yang lain. Netizen mengedukasi netizen lain bahwa fungsi vaksin memang bukan untuk membuat seseorang menjadi kebal terhadap Covid-19, tapi sebagai bekal perlindungan agar gejala Covid-19 yang dialami tidak separah sebelum

menerima vaksin. Topik ini diperbincangkan sebagai counter opini terhadap netizen yang bersentimen negatif pada vaksin.

3. Topik terpopuler yang ditemui dari percakapan netizen adalah harapan dan kepercayaan masyarakat menanggapi kebijakan vaksinasi pemerintah. Netizen juga memperbincangkan kegiatan vaksinasi pada dosis tertentu yang telah diikuti oleh sesama netizen dan informasi

No	Perkiraan Tema	Kata-kata yang mewakili
1	Harapan Masyarakat	'udah', 'booster', 'warganya', 'berharap', 'now', 'bertahan', 'semangat', 'kekebalan', 'percaya', 'persentase'
2	Kebijakan Pemerintah	'pemerintah', 'memperluas', 'provokator', 'guys', 'efektifitas', 'rantai', 'karna', 'dan' anak', 'mari'
3	Dosis Vaksin	'emang', 'vaksin', 'covid', 'dosis', 'dosis', 'tingkat', 'efektifitas', 'nya', 'aja', 'si', 'vaksin', 'covid', 'dosis', 'netral', 'nan', 'positif'
4	Penanganan Covid	'jumpai', 'dki', 'disediakan', 'warganya', 'kunci', 'cibinong', 'alasan', 'penyebarannya', 'kenali', 'menyambut'
5	Peran Komnas	'hasil', 'lengah', 'komnas', 'konsumsi', 'ya', 'mati', 'jalannya', 'emang', 'jumat', 'peran'
6	Gelombang Vaksin di Tiongkok	'ingat', 'tongkok', 'terbukti', 'dampak', 'menjaga', 'hidup', 'lapangan' 'tambahan', 'toko', 'divaksin'
7	Peran Vaksin	'bangsa', 'tameng', 'benteng', 'prosedurnya', 'kemungkinan', 'menggunakan', 'lengah', 'motif', 'aman', 'taat'
8	Manfaat Vaksin	'harian', 'saya', 'khofifah', 'sistem', 'manfaatnya', 'tracking', 'ketularan', 'jenis', 'namanya', 'sumatera'
9	Vaksinasi Anak	'kelompok', 'timbangan', 'stay', 'loh', 'fyi', 'dki', 'kerjakan', 'papua', 'upaya'
10	Peran Vaksin	'bangsa', 'tameng', 'benteng', 'prosedurnya', 'kemungkinan', 'menggunakan', 'lengah', 'motif', 'aman', 'taat'
11	Manfaat Vaksin	'harian', 'saya', 'khofifah', 'sistem', 'manfaatnya', 'tracking', 'ketularan', 'jenis', 'namanya', 'sumatera'
12	Info Vaksinasi Daerah	'kelompok', 'timbangan', 'stay', 'loh', 'fyi', 'dki', 'kerjakan', 'papua', 'upaya'

yang berkaitan dengannya.

REFERENSI

- [1] Kementerian Kesehatan Republik Indonesia, *Buku Saku Tanya Jawab Seputar Vaksinasi Covid 19*, 1st ed. Jakarta: Kementerian Kesehatan Republik Indonesia, 2021.
- [2] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, vol. 13, no. 3, p. 319, Sep. 1989, doi: 10.2307/249008.
- [3] A. Szabolcsi, "Positive Polarity – Negative Polarity," *Nat. Lang. Linguist. Theory*, vol. 22, no. 2, pp. 409–452, May 2004, doi: 10.1023/B:NALA.0000015791.00288.43.
- [4] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web - WWW '10*, 2010, p. 591. doi: 10.1145/1772690.1772751.
- [5] G. Valkanas, A. Saravanou, and D. Gunopulos, "A Faceted Crawler for the Twitter Service," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8787, 2014, pp. 178–188. doi: 10.1007/978-3-319-11746-1_13.
- [6] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, Nov. 2021, doi: 10.1016/j.jjime.2021.100019.
- [7] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web," in *Proceedings of the 14th International Conference on World Wide Web*, 2005.
- [8] Fajar Fathur Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Indones. Heal. Inf. Manag. J.*, vol. 8, no. 2, 2020.
- [9] J. C. Campbell, A. Hindle, and E. Stroulia, "Latent Dirichlet Allocation," in *The Art and Science of Analyzing Software Data*, Elsevier, 2015, pp. 139–159. doi: 10.1016/B978-0-12-411519-4.00006-9.