

**Jurnal Politeknik Caltex Riau**Terbit Online pada laman <https://jurnal.pcr.ac.id/index.php/jkt/>

| e- ISSN : 2460-5255 (Online) | p- ISSN : 2443-4159 (Print) |

Analisis Sentimen Pendapat Masyarakat Mengenai Vaksin Covid-19 Pada Media Sosial Twitter dengan Robustly Optimized BERT Pretraining Approach

Muhammad Mahrus Zain¹, Rizky Nathamael Simbolon², Harlem Sulung³ dan Zaidan Anwar⁴¹Politeknik Caltex Riau, Jurusan Teknologi Informasi, email: mahrus@pcr.ac.id²Politeknik Caltex Riau, Jurusan Teknologi Informasi, email: rizky20si@mahasiswa.pcr.ac.id³Politeknik Caltex Riau, Jurusan Teknologi Informasi, email: harlem20ti@mahasiswa.pcr.ac.id⁴Politeknik Caltex Riau, Jurusan Teknologi Informasi, email: zaidan20ti@mahasiswa.pcr.ac.id

[1] Abstrak

Pada tanggal 11 Maret 2020, World Health Organization (WHO) resmi menetapkan COVID-19 sebagai pandemi global. Penyebaran virus ini sudah ada pada tahun 2019 lalu di kota Wuhan, China. Pemerintah resmi menetapkan Peraturan Presiden (PERPRES) tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi Coronavirus Disease. Rencana kegiatan vaksinasi tersebut juga haruslah mempertimbangkan berbagai masukan, di antaranya adalah dengan melihat bagaimana respon dan opini masyarakat terhadap wacana vaksinasi tersebut. Dengan memanfaatkan data dari media sosial twitter, penelitian ini bertujuan untuk menganalisis respon masyarakat terhadap wacana vaksinasi dengan cara mengklasifikasikan respon tersebut ke dalam respon positif dan negatif. Selanjutnya juga akan dilakukan pengelompokan opini masyarakat menggunakan model pre-trained Indonesian RoBERTa Base Sentiment Classifier untuk mengetahui sentimen dari topik vaksinasi Covid-19 yang dibahas oleh masyarakat. Hasil analisis menunjukkan bahwa masyarakat lebih banyak memberikan respon negatif terhadap wacana tersebut (24,7%) dibandingkan dengan respon positifnya (5,7%) dengan sisa respon netral (69,6%). Kata-kata bersentimen yang paling sering muncul juga mengindikasikan lebih banyak kata yang bersentimen negatif dibandingkan dengan kata yang bersentimen positif. Rata-rata hasil akurasi prediksi penerapan model pre-trainer pada label positif adalah 84%, Netral 97% dan Negatif 93%.

Kata kunci: COVID-19, vaksin, sentimen, twitter, NLP

[2] Abstract

On March 11, 2020, the World Health Organization (WHO) officially declared COVID-19 a global pandemic. The spread of this virus already existed in 2019 in the city of Wuhan, China. The government officially stipulates a Presidential Regulation (PERPRES) on Vaccine Procurement and Vaccination Implementation in the Context of Overcoming the Coronavirus Disease Pandemic. The vaccination activity plan must also consider various inputs, among them is by looking at how the response and public opinion to the vaccination discourse. By utilizing data from Twitter social media, this study aims to analyze the public's response to the vaccination discourse by classifying the response into positive and negative responses. Furthermore, public opinion grouping will also be carried out using the pre-trained Indonesian RoBERTa Base

Sentiment Classifier model to find out the sentiments of the Covid-19 vaccination topic discussed by the community. The results of the analysis showed that the public gave more negative responses to the discourse (24.7%) compared to positive responses (5.7%) with the remaining neutral responses (69.6%). Sentiment words that occur most often also indicate more words with negative sentiments than words with positive sentiments. The average results of the prediction accuracy of the application of the pre-trainer model on the positive label are 84%, Neutral 97% and Negative 93%.

Keywords: COVID-19, vaccine, sentiment, twitter, NLP

1. Pendahuluan

Pada tanggal 11 Maret 2020, World Health Organization (WHO) resmi menetapkan COVID-19 sebagai pandemi global. Penyebaran virus ini sudah ada pada tahun 2019 lalu di kota Wuhan, China [1]. Berdasarkan data dari covid19.go.id, pada bulan juli 2021 mengalami peningkatan kasus yang cukup tinggi. Pemerintah Indonesia banyak menerapkan berbagai cara dalam menghambat persebaran kasus covid-19, salah satu nya adalah vaksinasi penduduk secara menyeluruh. Rencana kegiatan vaksinasi tersebut haruslah mempertimbangkan segala aspek, mulai dari aspek kelayakan vaksin yang akan digunakan[2], risiko pasca pemakaian, sampai tahapan dan prosedur dari pemberian vaksin hingga nantinya sampai ke masyarakat [7]. Pengumpulan pendapat mengenai vaksin melalui media sosial dinilai dapat memberikan efisiensi dalam segala hal apabila dibandingkan dengan harus melakukan survey tradisional. Kegiatan menganalisis respon dan opini masyarakat menggunakan data yang bersumber dari media sosial twitter juga telah banyak dilakukan pada penelitian-penelitian sebelumnya, misalnya penelitian yang dilakukan untuk melihat opini masyarakat terhadap kebijakan ganjil genap di India[14] dan melihat bagaimana opini masyarakat terhadap pelayanan LRT di Los Angeles [15]. Berdasarkan latar belakang tersebut, maka penelitian ini ingin melihat bagaimana respon dan opini masyarakat Indonesia terhadap vaksin COVID-19 dengan menggunakan data yang bersumber dari media sosial twitter.

Oleh karena itu penulis melakukan penelitian untuk menganalisa pendapat masyarakat mengenai vaksin untuk mendukung program percepatan vaksin dengan mengklasifikasikan pendapat dengan kategori positif, negatif dan netral pada penggunaan media sosial twitter dengan menggunakan metode Robustly Optimized BERT Pretraining Approach (RoBERTa). Model RoBERTa yang digunakan adalah Indonesian RoBERTa Base Sentiment Classifier untuk melakukan pelabelan sentimen pendapat masyarakat indonesia pada media sosial twitter. Data yang sudah diberikan label akan dijadikan data latih dan data uji untuk penerapan metode Long Short-Term Memory untuk melakukan pengujian dan analisis presisi. Pengolahan data menggunakan bahasa pemrograman Python dan menggunakan Google Collaboratory..

2. Studi Literatur

a. Penelitian Terkait

Opinion Mining merupakan proses untuk melakukan ekstraksi atribut dan komponen objek menentukan apakah sebuah pendapat atau opini itu positif, negatif atau netral [11]. Opinion mining merupakan salah satu cabang penelitian text mining, yang merupakan riset komputasional dari opini, sentimen, dan emosi yang diekspresikan. Text mining merupakan suatu teknik untuk melakukan ekstraksi informasi yang berguna dari sumber data dalam bentuk text untuk mendapatkan pengetahuan dan eksplorasi pola yang menarik. Text mining cenderung mengarah pada bidang penelitian data mining. Oleh karena itu, text mining berada pada arsitektur yang

sama dengan data mining [12]. RoBERTa merupakan penyempurnaan model BERT asli bersama dengan manipulasi data dan input. RoBERTa sangat bergantung pada jumlah teks. Jika semakin banyak data maka akan semakin meningkat kinerjanya.

Sarkasme adalah gaya bahasa yang memiliki polaritas negatif serta sulit dideteksi secara otomatis oleh mesin komputer, bahkan oleh manusia[6]. Dewi Hernikawati[10] menggunakan metode Lexicon Based, yang merupakan metode yang sederhana, layak, dan praktis untuk analisis sentimen. Matulatuwa, Sedyono, dan Iriani [15] Data yang bisa digunakan berasal dari media sosial seperti Twitter, Facebook, dan media sosial lain mengenai opini suatu produk atau layanan jasa

Terkait dengan metode Latent Dirichlet Allocation (LDA), Fajar Fathur Rachman dan Setia Pramana[3] menyatakan bahwa LDA merupakan metode topic modeling yang paling populer saat ini. LDA muncul sebagai salah satu metode yang melakukan analisis pada dokumen berukuran sangat besar. LDA dapat digunakan untuk meringkas, melakukan klusterisasi, menghubungkan juga memproses data yang besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen.

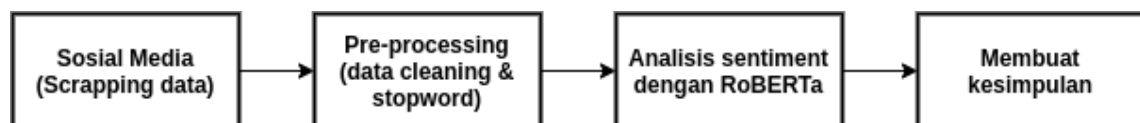
Kemudian menurut Ahmad Harun dan Dea Putri Ananda[5], Naïve Bayes Classifier (NBC) adalah algoritma klasifikasi berdasarkan teorema Bayes dengan asumsi independen yang digunakan untuk memprediksi data seakurat mungkin. Tabel berikut adalah rangkuman dari penelitian yang dilakukan oleh [10][3][5]:

Tabel 1. Penelitian yang pernah dilakukan oleh [10][3][5]

Peneliti	Data Set	Metode Pendeteksi Sarkasme	Hasil Pengujian Metode
Dewi Hernikawati	Data teks dari Twitter dengan kata kunci “vaksin Sinovac” dan “Sinovac Vaccine”	<ul style="list-style-type: none"> • Crawling data API Twitter dengan <i>library</i> Tweepy pada Python • Membersihkan data menggunakan <i>Regular Expression</i> dengan <i>library</i> re pada Python • Analisis sentimen dengan menggunakan metode <i>Lexicon Based</i> 	<ul style="list-style-type: none"> • <i>Positive Tweets</i> 35.4% di Indonesia dan 21.6 % di Dunia • <i>Negative Tweets</i> 27.0% di Indonesia dan 9.0% di Dunia • <i>Neutral Tweets</i> 37.6% di Indonesia dan 69.4 % di Dunia
Fajar Fathur Rachman, Setia Pramana	Data teks dari Twitter dengan kata kunci “Vaksin Covid” dan “Vaksin Corona”	<ul style="list-style-type: none"> • Analisis sentimen dengan menggunakan metode <i>Lexicon Based</i> atau berbasis kamus positif-negatif. • Pengelompokan opini masyarakat dengan metode <i>Latent Dirichlet Allocation</i> (LDA) 	<ul style="list-style-type: none"> • Sentimen Positif sebanyak 29,6% • Sentimen Netral sebanyak 46,8% • Sentimen Negatif sebanyak 23,6%
Ahmad Harun, Dea Putri Ananda	Data teks dari Facebook Page Kementrian Kesehatan	Naïve Bayes Classifier dan Decision Tree	Naïve Bayes Classifier dengan nilai akurasi 100.00% dan Decision

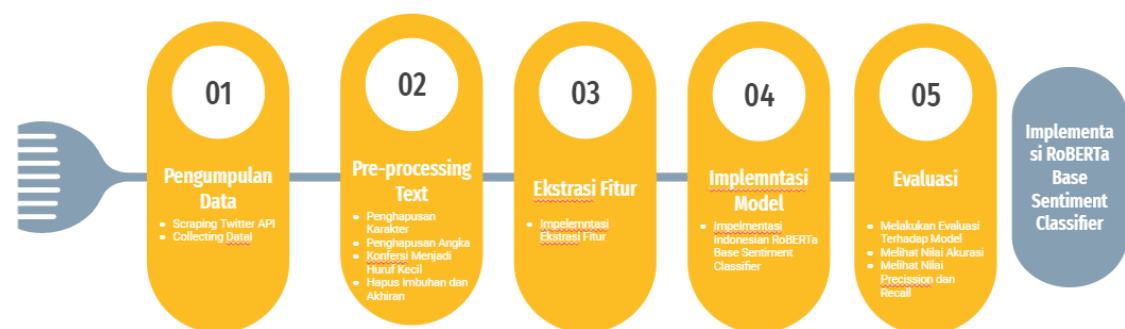
Republik Indonesia	Tree dengan nilai akurasi 50.39%
-----------------------	-------------------------------------

Pada penelitian ini akan melakukan penelitian dengan tahapan yang ada pada Gambar 1. Tahapan pertama adalah pengumpulan data dengan cara mengumpulkan data berupa opini dan respons dari masyarakat yang diambil dari media sosial twitter. Selanjutnya tahapan Pre-processing atau proses untuk mempersiapkan data mentah sebelum dilakukan proses pelabelan dengan metode RoBERTa. Pada tahapan sentimen analisis akan dilakukan pelabelan dengan menerapkan model pre-trained Indonesian RoBERTa Base Sentiment Classifier. Indonesian RoBERTa merupakan model berbasis RoBERTa yang telah dilatih menggunakan 522MB Wikipedia bahasa Indonesia menggunakan Masked Language Modeling (MLM).



Gambar 1. Tahapan Sistem

3. Metode yang Digunakan



Gambar 2 –Implementasi Metode

a. Pengumpulan Data

Teknik yang digunakan dalam pengumpulan data menggunakan teknik web scraping pada twitter. Kegiatan web scraping ini menggunakan 'twint'. Twint adalah tools scraping twitter yang ditulis dengan bahasa Python yang mengizinkan scraping tweet tanpa harus menggunakan API key. Kelebihan dari tools twint ini adalah dapat mengambil hampir semua tweet, dapat digunakan secara anonim tanpa mendaftar akun twitter dan tidak ada batasan tarif. Kata kunci yang digunakan untuk menyaring respon dan dan opini masyarakat menggunakan kata kunci "vaksin covid" dengan batasan tweets yaitu 109202 tweets.

b. Pre-processing Text

Data pre-processing adalah sebuah tahapan yang sangat penting di mana dalam melakukan proses klasifikasi data teks. Tujuan dilakukannya data pre-processing yaitu untuk menghilangkan noise atau data yang tidak dibutuhkan seperti emoji, url, angka, dan untuk menyeragamkan bentuk data kata sehingga data kata dapat menjadi lebih bersih sebelum diolah pada proses selanjutnya [9]. Dari data yang didapatkan masih banyak ditemukan kata-kata yang tidak berstruktur sehingga dibutuhkan pemrosesan atau pembersihan data yang dilakukan yaitu case folding untuk Mengubah semua huruf kapital menjadi huruf kecil. Kemudian dilakukan data

Cleaning: Menghapus data yang terdandan, menghilangkan symbol, url, angka, string ASCII, emote, singkatan['yg', 'dg', 'tidak', dll], dan tanda baca seperti seperti ["TM", "C", "SM"];

c. Indonesian RoBERTa Base Sentiment Classifier

Data yang telah selesai dibersihkan dilakukan text processing dan labelling dengan menerapkan model pretrained dengan model Indonesian RoBERTa Base Sentiment Classifier[4]. Klasifikasi RoBERTa Base Sentiment Indonesia adalah sebuah model klasifikasi sentimen teks berdasarkan model RoBERTa. Model tersebut awalnya adalah model RoBERTa Base Indonesia yang telah dilatih sebelumnya, yang kemudian disempurnakan pada dataset SmSA indonlu yang terdiri dari komentar dan ulasan berbahasa Indonesia. Setelah model dilatih, pencapaian akurasi model evaluasi 94,36% dan F1-makro 92,42%. Pada set uji benchmark, model mencapai akurasi 93,2% dan F1-makro 91,02%. [14]

4. Hasil dan Pengujian

a. Scraping

Hasil yang ditemukan dari teknik *web scraping*, di mana kata kunci yang digunakan adalah "vaksin covid" dengan jumlah data sebanyak 109202 cuitan. Penyaringan yang dilakukan berupa pembatasan lokasi yaitu negara indonesia dan batasan datanya hanya 109202 data. Setelah proses scraping dilakukan, kemudian data mentah ini disimpan dan diperoleh 36 atribut yang terdiri dari id(id cuitan), conversation_id(id percakapan), create_at(waktu cuitan) yang kemudian dipecahkan menjadi date(tanggal) dan time(jam) dan masih banyak atribut lainnya. Tetapi dari atribut tersebut yang memiliki peran penting dalam *text mining* adalah atribut "tweet" yang mana atribut ini merupakan data dari cuitan pengguna *twitter* yang jikalau diolah, maka menghasilkan banyak pengetahuan. Setelah kegiatan *scraping data*, jumlah *tweets* yang siap untuk proses *Pre-processing* dan dianalisis yaitu sebanyak 109202 data

b. Pre-processing

Tahap *pre-processing* atau persiapan data merupakan tahapan yang paling penting dan krusial dalam sebuah penelitian dengan menggunakan metode *text mining*. Data mentah dari hasil kegiatan *text mining* harus disiapkan terlebih dahulu sesuai dengan kebutuhan analisis sehingga dapat menghasilkan hasil analisis yang baik. Misalnya pada penelitian ini salah satu tahapan dalam kegiatan *pre-processing* adalah tahapan mengubah semua ukuran font menjadi huruf kecil (*lowercase*). Hal ini penting dilakukan karena sebuah kata yang terdiri dari huruf yang sama akan dibaca oleh sistem sebagai kata yang berbeda jika huruf tersebut memiliki ukuran huruf yang berbeda. Misalnya seperti kata 'Ceria' dan 'ceria'. Jika hal ini dibiarkan maka akan berdampak pada waktu komputer untuk memproses data. Namun yang terpenting, hasil yang akan diperoleh menjadi ambigu atau tidak jelas.

Pada penelitian ini, penulis mengelola data yang didapat melalui scraping sebelumnya, dengan contoh perbedaan data sebelum & sesudah melalui *pre-processing* sebagai berikut:

Tabel 2 – Perbedaan Data Sebelum dan Sesudah Pre-processing

Sebelum	Sesudah
berikut adalah vaksin yang diakui oleh pemerin...	vaksin diakui pemerintah saudi arabia jamaah u...

temenku send ini, after 3hari aku udh divaksin...

@hibooraan ih iya sebagian orang pada gak mau d...

@mjazryjaafar @admiralkunkka @amirhimself jawa...

rt jokowi "sebanyak 21,2 juta dosis bahan baku...

temenku send after divaksin puskesmas didaerah...

ih iya orang daerah gue gila banget percaya co...

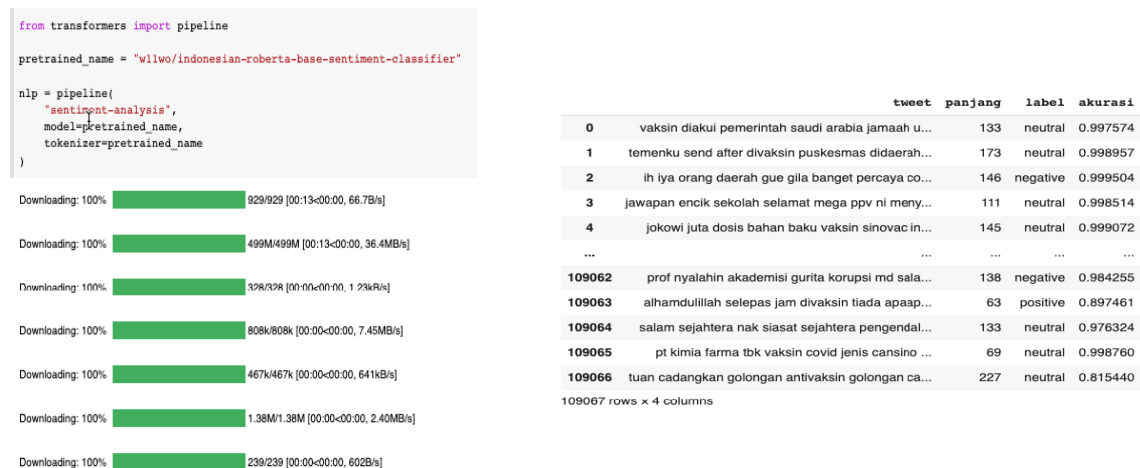
jawapan encik sekolah selamat mega ppv ni meny...

jokowi juta dosis bahan baku vaksin sinovac in...

Terlihat pada perbandingan di atas, bahwa data pada bagian sebelum dan sesudah memiliki beberapa perbedaan yang telah disaring meliputi: penghapusan regular expression["tabulasi", "baris baru", ...], penghapusan karakter Non-ASCII["Emot", "Huruf China", ...], penghapusan huruf, tanda baca, *trailing-spacing* & spasi duplikat, juga penghapusan *stopword*. Tahapan ini dilakukan dengan menggunakan beberapa *library* bawaan dari Python *build* itu sendiri salah satunya *library regex*, dan juga *library external NLTK*.

c. Indonesian RoBERTa Base Sentiment Classifier

Hasil dari data yang sudah dibersihkan akan diimplementasikan ke dalam model pretrained *Indonesian RoBERTa Base Sentiment Classifier* untuk menghasilkan label dan akurasi dari prediksi label. Di bawah ini adalah hasil implementasi dan hasil implementasi *Indonesian RoBERTa Base Sentiment Classifier*. Proses prediksi memakan waktu kurang lebih 2 jam 23 menit untuk melakukan klasifikasi sentimen terhadap 109.202 *tweet*.



Gambar 3 – Implementasi Indonesian RoBERTa BSC dan Hasil yang didapatkan

d. Visualisasi Hasil

Nilai sentimen untuk setiap *tweets* pada penelitian ini diperoleh dari perhitungan jumlah kata positif, negatif, dan netral yang terdapat pada suatu *tweets*. Dapat dilihat dari tabel berikut bahwa respon dari mayoritas masyarakat Indonesia terhadap Vaksin Covid-19 adalah netral, yakni sebesar 69,6%. Hal tersebut mengindikasikan bahwa masyarakat cenderung memberikan respon yang bersentimen netral dibandingkan dengan respon yang bersentimen positif maupun negatif. Kemudian 24,7% masyarakat memberikan respon yang bersentimen negatif, yang di

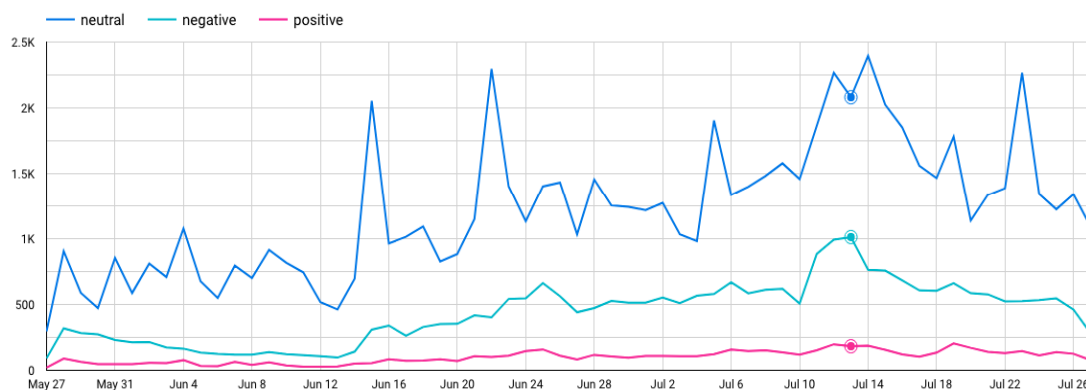
mana respon tersebut memiliki kesenjangan yang cukup jauh bila dibandingkan dengan respon yang bersentimen positif. Masyarakat yang berpandangan negatif ini, cenderung menganggap remeh Covid-19 dan abai terhadap protokol kesehatan. Karena masih banyaknya masyarakat yang abai terhadap Covid-19 dan tidak percaya dengan Vaksin Covid-19, pengendalian Covid-19 di Indonesia menjadi sulit dan cenderung tidak terkendali. Hal ini yang menyebabkan rekor atas penambahan kasus baru (per tanggal 14 Juli 2021) sebanyak 54.517 kasus baru.

Berikut adalah tabel Jumlah dan persentase respon publik terhadap vaksin COVID-19 berdasarkan jenis sentimennya:

Tabel 3 – Jumlah dan Persentase Respon Publik terhadap Vaksin Covid-19

Sentimen	Jumlah	Persentase
Positif	6250	5,7%
Netral	75883	69,6%
Negatif	26934	24,7%

Berikut adalah grafik frekuensi sentimen atas respon masyarakat terhadap Vaksin Covid-19 dari hari ke hari (per tanggal 27 Mei 2021 sampai 26 Juli 2021):



Gambar 4 – Grafik Frekuensi Sentimen

Pada bulan Mei-Juni 2021, dapat dilihat bahwa respon negatif masyarakat terhadap Vaksin Covid-19 cenderung meningkat. Hal ini juga berbanding lurus dengan penambahan kasus positif harian yang kian meningkat. Masyarakat yang bersentimen negatif terhadap Vaksin Covid-19, cenderung abai dengan Virus Covid-19 dan tidak jarang mereka mengabaikan protokol kesehatan. Hal ini tentu dapat memperkeruh kondisi pandemi di Indonesia. Tanggal 14 Juli 2021 merupakan titik balik bagi mereka yang abai terhadap Covid-19.

Per 14 Juli 2021, terdapat lonjakan atas respon masyarakat terhadap Vaksin Covid-19 yang signifikan. Hal tersebut dapat terjadi karena pada tanggal 14 Juli 2021, kasus positif Covid-19 di Indonesia mencatatkan rekor atas penambahan kasus baru sebanyak 54.517. Lonjakan kasus ini tentu dapat membuat kepercayaan masyarakat terhadap Vaksin Covid-19 yang sebelumnya

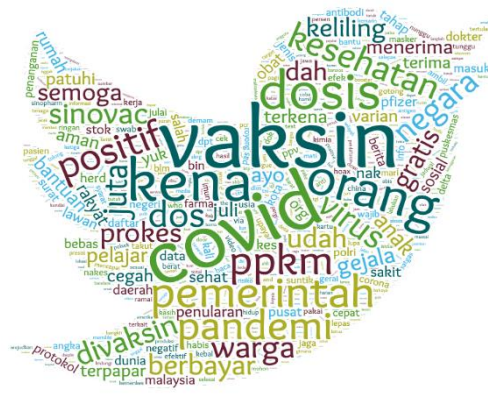
berpandangan negatif terhadap Vaksin Covid-19, berangsur menjadi berpandangan positif maupun netral, mengingat kondisi Covid-19 yang sudah semakin parah, dan munculnya kluster keluarga. Setelah beberapa hari berselang, kasus harian positif Covid-19 berangsur turun karena kepercayaan masyarakat terhadap Vaksin Covid-19 yang meningkat.



Gambar 5 – Word Cloud Negatif



Gambar 6 – Word Cloud Positif



Gambar 7 – Word Cloud Netral

Word Cloud di atas menggambarkan kata-kata mana yang biasanya digunakan sehubungan dengan setiap sentimen dari *tweet* yang menyebutkan *Tweet* positif, negatif, dan netral. *Tweet* positif tersebut tampaknya berasal dari orang yang baru saja atau telah menerima vaksin pertama mereka, ada juga yang berterima kasih atas pekerjaan yang dilakukan para ilmuwan dan petugas kesehatan. Sedangkan *tweet* negatif tampaknya berasal dari orang-orang yang menderita efek samping, reaksi merugikan, atau ketidaknyamanan umum setelah menerima vaksin. Kata-kata yang digunakan dalam *tweet* netral yang berkaitan dengan vaksin Covid-19. Beberapa dari kata-kata ini termasuk "kesehatan", "sembuh", "suntik", dan lain-lainnya.

5. Hasil

Hasil analisis pengujian atas terhadap klasifikasi sentimen dari model yang digunakan, telah diukur nilai akurasi pada setiap label yang diprediksi oleh model Indonesian RoBERTa Base Sentiment Classifier. Dari jumlah data yang digunakan sebanyak 109202 data, yang telah diprediksi menghasilkan 6.250 sentimen positif, 75.883 sentimen netral dan 26.934 negatif. Rata-rata hasil akurasi prediksi bisa dilihat pada tabel di bawah ini. Tabel 4 di bawah ini merupakan rata-rata hasil akurasi yang diprediksi oleh model pre-trained.

Tabel 4 – Rata-rata Hasil Akurasi Model Pre-trained

Sentimen	Rata-rata Hasil Akurasi
Positif	0.84
Netral	0.97
Negatif	0.93

6. Kesimpulan

Ringkasan dari penelitian analisis sentimen dengan mengimplementasikan pre-trained Indonesian RoBERTa Base Sentiment Classifier di mana data diklasifikasikan ke dalam 3 kelas yang berbeda telah dilakukan. Pada penelitian ini hasil prediksi yang dilakukan memiliki akurasi yang baik dengan rata-rata akurasi keseluruhan data sebesar 95%. Sedangkan rata-rata hasil akurasi prediksi pada label positif adalah 84%, Netral 97% dan Negatif 93%.

Daftar Pustaka

- [1] S. S. Al Jameel et al., "A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 218, 2020, doi: 10.3390/ijerph18010218
- [2] Liu C, Zhou Q, Li Y, Garner L V, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. 2020
- [3] F. F. 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, pp. 100– 109, 2020
- [4] Wongso, W. (n.d.). *w11wo/indonesian-roberta-base-sentiment-classifier Hugging Face*. Retrieved from huggingface.co: <https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier>
- [5] A. Harun and D. P. Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree". 2021. Retrieved from <https://journal.irpi.or.id/index.php/malcom/article/view/63/31>
- [6] S. Christina, "Sarcasm in Sentiment Analysis of Indonesian Text: A Literature Review". 2019. Retrieved from <https://media.neliti.com/media/publications/294966-sarcasm-in-sentiment-analysis-of-indones-bd55d2e2.pdf>
- [7] Peraturan Presiden (PERPRES) Nomor 99 Tahun 2020
- [8] K. Rajeswari and P. Shanthibala, "Recognition of Sarcastic Emotions of Individuals on Social Network," *Int. J. Pure Appl. Math.*, vol. 118, no. 7, pp. 253-259, 2018

- [9] E. Lunando and A. Purwarianti, "Indonesian Social Media Sentiment Analysis with Sarcasm Detection," *J. Sarj, Inst. Teknol. Bandung Bid. Tek. Elektro dan Inform.*, 2013.
- [10] D. Hernikawati, "Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis," *J. IPTEK-KOM.*, 2021
- [11] Bing Liu.(2010). Opinion Mining. Departemen of Computer Science, University of Illinois at Chicago
- [12] Ronen Feldman. (2008). Applied Text Mining. Information Systems Department School Of Business Administration Hebrew University, Jerusalem
- [13] RoBERTa: Pendekatan Pra-pelatihan BERT yang Dioptimalkan dengan Kua
- [14] <https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier>
- [15] Matulatuwa, Febrilien Matresya, Eko Sedyono, and Ade Iriani. 2017. "Text Mining Dengan Metode Lexicon Based Untuk Sentiment Analysis Pelayanan PT. Pos Indonesia Melalui Media Sosial Twitter." *Jurnal Masyarakat Informatika Indonesia* 2 (3): 52–65