

# ANALISIS SENTIMEN PERKEMBANGAN KASUS COVID-19 PADA KOMENTAR FACEBOOK

Irsyad Arif Mashudi<sup>1</sup>, Sofyan Noor Arief<sup>2</sup>

Jurusan Teknologi Informasi, Politeknik Negeri Malang  
email: <sup>1</sup>irsyad.arif@polinema.ac.id, <sup>2</sup>sofyan@polinema.ac.id

(Artikel diterima: Desember 2020, diterima untuk terbit: Januari 2021)

**Abstrak** – Masyarakat banyak yang menggunakan sosial media sebagai sumber berita. Diantaranya, Facebook masih menjadi media sosial yang paling banyak digunakan masyarakat. Hal ini berlaku juga dengan berita terkait COVID-19. Sebagian besar masyarakat Indonesia mengikuti perkembangan berita tentang COVID-19 melalui Facebook. Dengan mengikuti di Facebook, masyarakat dapat menanggapi dan berkomentar terhadap isu-isu terkait COVID-19. Saat ini, sudah banyak tanggapan yang dapat dilihat terkait COVID-19. Penelitian ini ditujukan untuk menganalisis sentimen komentar warganet terkait perkembangan kasus COVID-19. COVID-19 masih akan berada di tengah masyarakat dalam waktu lama. Karena itu, kewaspadaan masyarakat juga perlu dijaga. Hal ini dilakukan agar dapat mempercepat berakhirnya pandemi ini. Tingkat kewaspadaan itu salah satunya tercermin dari komentar masyarakat terhadap perkembangan COVID-19. Komentar yang ada dapat dilakukan analisis sentimen yang mencerminkan kewaspadaan masyarakat terhadap kasus COVID-19. Empat model telah dibuat untuk melakukan hal ini. Hasilnya, keempat model menghasilkan analisis sentimen yang berbeda-beda. Model Naïve Bayes dengan tingkat akurasi tertinggi lebih banyak mendeteksi sentimen negatif. Sebaliknya, model Random Forest yang lebih netral lebih banyak mendeteksi sentimen positif. Hasil klasifikasi ini dapat digunakan pemerintah dalam mengambil kebijakan terkait COVID-19

**Kata kunci:** analisis sentimen, COVID-19, media sosial, SVM, Logistic Regression, Random Forest, Naïve Bayes

**Abstract** – Many people use social media as a news source. Among them, Facebook is still the most widely used social media community. This also applies to news related to COVID-19. Most Indonesians follow news developments about COVID-19 through Facebook. By following on Facebook, people can respond and comment on issues related to COVID-19. Currently, there are many responses that can be seen regarding COVID-19. This study aims to analyze the sentiments of netizens' comments regarding the development of the COVID-19 case. COVID-19 will still be in the midst of society for a long time. Therefore, public awareness also needs to be maintained. This is done in order to hasten the end of this pandemic. This level of vigilance is reflected in the public's comments on the development of COVID-19. Existing comments can be carried out by sentiment analysis that reflects public awareness of COVID-19 cases. Four models have been created to do this. As a result, the four models produce different sentiment analyzes. The Naïve Bayes model with the highest level of accuracy detects more negative sentiments. In contrast, the more neutral Random Forest model detects more positive sentiment. The results of this classification can be used by the government in making policies related to COVID-19.

**Keywords:** sentiment analysis, COVID-19, social media, SVM, Logistic Regression, Random Forest, Naïve Bayes

## I. PENDAHULUAN

COVID-19 merupakan penyakit pernafasan yang sampai saat ini pengobatannya belum ditemukan [1]. Penyakit ini memiliki gejala seperti influenza. Walaupun begitu, karena belum ada pengobatan yang efektif, tingkat mortalitas penyakit ini jauh lebih mematikan. Saat ini, tindakan yang paling efektif adalah upaya preventif agar tidak terkena penyakit ini.

Di Indonesia, COVID-19 terdeteksi sejak awal Maret 2020. Sejak itu, kebiasaan masyarakat Indonesia terpaksa diubah demi menekan perkembangan penyakit ini. Walau angka kematian penyakit ini terus menurun, belum ada tanda-tanda bahwa penyakit ini akan hilang dalam waktu dekat [2]. Penyakit ini telah mengubah tatanan masyarakat dan perkembangannya terus diikuti masyarakat umum melalui media-media yang dapat diakses secara jamak.

Banyak masyarakat Indonesia yang mendapatkan akses berita COVID-19 melalui media sosial. Pada media sosial, masyarakat dapat memberikan berbagai tanggapan terhadap berita mengenai penyakit ini. Tanggapan ini dapat berupa tanggapan yang positif atau negatif. Tanggapan-tanggapan ini dapat dianalisa untuk memprediksi kepedulian masyarakat terhadap kasus COVID-19.

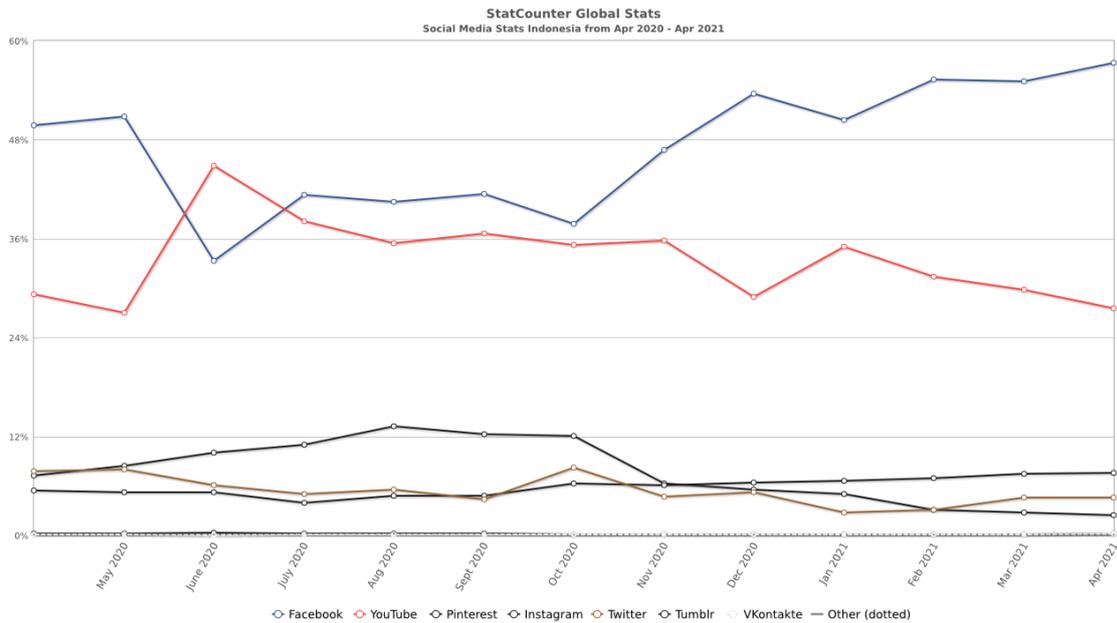
Pada saat artikel ini ditulis, COVID-19 sudah berlangsung lebih dari setahun. Sebagian elemen masyarakat pun sudah mulai terbiasa dengan adanya COVID-19 [3].

Padahal, wabah penyakit belum berakhir. Masih belum ada obat untuk COVID-19. Kewaspadaan harus tetap dijaga agar COVID-19 yang sudah turun tidak naik kembali.

Pemerintah perlu menjaga kewaspadaan masyarakat terhadap COVID-19 agar dampak dari penyakit ini tidak berlarut-larut. Tindakan yang tepat perlu ditentukan agar kewaspadaan masyarakat tetap terjaga. Salah satu aspek yang dapat menentukan tepat tidaknya tindakan yang diambil adalah sentimen masyarakat terhadap perkembangan kasus COVID-19. Jika banyak sentimen negatif, maka pemerintah perlu menenangkan masyarakat. Sebaliknya, jika banyak sentimen positif, maka pemerintah perlu mengingatkan bahwa COVID-19 belum berakhir.

Analisis sentimen adalah suatu metode untuk mempelajari sentimen, emosi, dan sikap dalam sebuah teks opini. Prinsip dasar dari analisis sentimen adalah metode untuk mengklasifikasikan polaritas dari teks tertentu dan menentukan apakah isinya lebih positif, netral atau negatif [4]. Penelitian ini bertujuan untuk mengetahui gambaran sentimen masyarakat Indonesia terhadap perkembangan kasus COVID-19. Dengan mengetahui sentimen masyarakat, pemerintah dapat memilih kebijakan penanganan pandemi dengan lebih efektif.

Telah banyak riset mengenai analisis sentimen terhadap berbagai topik yang berkaitan dengan COVID-19. Beberapa berkaitan dengan tanggapan masyarakat terhadap kebijakan pemerintah [5], [6]. Ada juga yang berkaitan dengan dampak COVID-19 pada aspek sosial dan ekonomi masyarakat [7]–



Gambar 1. Marketshare sosial media di Indonesia [12]

[9]. Paling banyak adalah analisis sentimen terhadap dampak COVID-19 pada kebiasaan masyarakat sebelum pandemi terjadi [10], [11].

Riset-riset tersebut dilakukan dengan berbagai macam data dan metode. Sebagian besar menggunakan data *tweet* dari media social Twitter [13]–[15]. Dan sebagian menggunakan menggunakan data campuran antara *tweet* dan data yang didapat dari berita daring.

Riset ini bertujuan untuk mencari sentimen dari data komentar yang ada. Data yang akan dicari sentimennya adalah data dari komentar Facebook. Facebook merupakan media sosial dengan pengguna terbanyak di Indonesia. Perbandingan pengguna antar media sosial di Indonesia dapat dilihat pada Gambar 1. Dengan menggunakan Facebook sebagai sumber data, diharapkan hasil yang didapat lebih mewakili pendapat dari masyarakat Indonesia.

## II. TINJAUAN PUSTAKA

Seperti yang telah disebutkan sebelumnya, sudah banyak riset mengenai analisis sentimen berkaitan dengan COVID-19. Riset-riset ini sebagian besar menggunakan data *tweet* sebagai sumber data utama. Riset-riset ini juga menggunakan berbagai macam metode dalam membuat model. Model yang dihasilkan memiliki akurasi yang bermacam-macam.

### A. Analisis Sentimen Berbahasa Indonesia

Untuk analisis sentimen berbahasa Indonesia telah ada *dataset* yang dapat digunakan untuk melatih model [16]. *Dataset* ini mencakup data utama sebesar 10.806 baris data berbahasa Indonesia yang diambil dari media sosial Twitter. Data ini telah dikategorikan ke dalam tiga label: positif, negatif, dan netral, selain itu, juga terdapat 454.559 baris data yang belum diproses. Setelah dilakukan pengujian menggunakan model analisis sentimen sederhana yang menggunakan algoritme SVM, KNN, dan SGD, terlihat bahwa nilai ketepatan dari data utama dan data pembanding seimbang pada masing-masing algoritme, dengan perbedaan

nilai ketepatan berkisar pada angka 4% sampai 12%. Ini membuktikan bahwa data yang disajikan sudah layak untuk digunakan dalam pemodelan analisis sentimen.

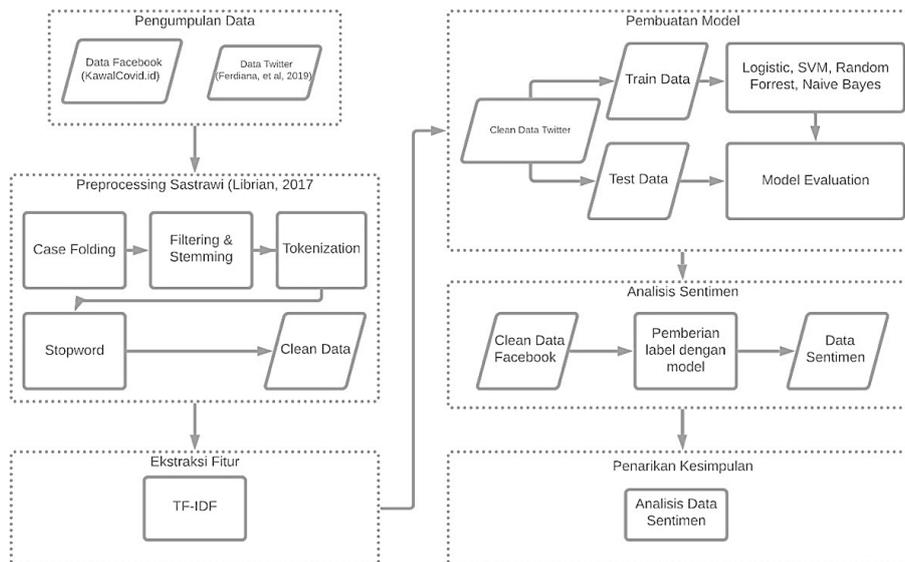
Prastyo dkk [5] mengembangkan algoritme pembelajaran mesin dan membandingkan SVM menggunakan Normalized Poly Kernel dan MNB untuk menentukan algoritme mana yang bekerja paling baik dalam melakukan analisis sentimen penanganan COVID-19 oleh pemerintah Indonesia. Dari segi ekonomi, masyarakat tampaknya setuju dengan kebijakan pemerintah dalam menangani COVID-19, namun masyarakat kurang puas dengan kinerja pemerintah secara umum. Analisis SVM pada sentimen aspek umum dengan menggunakan dataset dua kelas memiliki kinerja tertinggi pada rata-rata akurasi, presisi, recall, dan f-measure masing-masing sebesar 82.00%, 82.24%, 82.01%, dan 81.84%.

Selain dengan metode SVM, analisis sentimen mengenai COVID-19 juga banyak dilakukan dengan metode yang lain. Salah satunya adalah menggunakan Naïve-Bayes [7]. Hasil penelitian menunjukkan metode Naive Bayes mampu mengklasifikasi dokumen *tweet* dengan akurasi 67% dan error rate sebesar 33%. Penelitian dilakukan dengan menggunakan tiga volume data berbeda dan menghasilkan selisih nilai akurasi sebesar 0,02. Hal ini menunjukkan performa yang stabil dari metode Naive Bayes untuk klasifikasi data *tweet* terkait dampak virus Corona.

Ada juga yang membandingkan SVM dan Logistic Regression dalam melakukan analisis sentimen [18]. Penelitian ini mengumpulkan data *twitter* orang-orang di seluruh India dan kemudian mengukur polaritasnya menggunakan NLP. Setelah itu, digunakan metode klasifikasi SVM dan analisis Regresi Logistik untuk mengklasifikasikan sentimen. Hasilnya didapatkan SVM memberikan akurasi sebesar 91.50%, sedangkan Regresi Logistik memberikan akurasi sebesar 87.75%.

## III. METODE PENELITIAN

Penelitian ini menggunakan metode sebagai berikut. Pertama adalah pengumpulan data. Selanjutnya adalah pengolahan data yang telah dikumpulkan. Setelah data diolah,



Gambar 2. Detail metodologi penelitian

baru dibuat model. Model akan digunakan untuk menganalisis sentimen masyarakat secara umum. Hasil analisis ditarik kesimpulan. Gambaran metode penelitian secara umum bisa dilihat pada Gambar 2.

Ada dua dataset yang digunakan pada penelitian ini. Yang pertama adalah dataset twitter untuk analisis sentimen berbahasa Indonesia [16]. Dataset ini digunakan untuk melatih model yang dibuat. Yang kedua adalah dataset yang dikumpulkan dari *scrapping* komentar facebook pada halaman grup kawalcovid.id [19]. Grup ini adalah grup sukarelawan yang bergerak untuk mengawal informasi terkait COVID 19 secara transparan dan berdasarkan bukti. Data komentar yang diambil adalah komentar terkait perkembangan kasus COVID 19 di Indonesia selama bulan Januari dan Februari. Dataset ini yang akan dianalisis menggunakan model yang dibuat untuk menentukan sentimen masyarakat indonesia terhadap perkembangan kasus COVID 19.

Selanjutnya adalah pengolahan data. Data diolah dengan *preprocessing* menggunakan librari Sastrawi [20] dan ekstraksi fitur menggunakan metode TF-IDF. Pada *preprocessing*, pertama dilakukan Case Folding untuk menghilangkan huruf kapital. Selanjutnya, dilakukan *filtering* dan *stemming*. *Filtering* dilakukan untuk menghilangkan data string seperti hyperlink (contoh: www.abc.com), simbol (contoh: @, #, \$), dan gambar. *Stemming* dilakukan dengan mengubah kata berimbuhan menjadi kata dasar. Setelah itu, dilakukan *tokenization*. *Tokenization* dilakukan untuk mengubah data kalimat menjadi kata. Setelah data kalimat menjadi kata, dilakukan *stopword removal* untuk menghilangkan kata-kata yang tidak penting.

Setelah dilakukan *preprocessing*, tahap selanjutnya adalah ekstraksi fitur. Ekstraksi fitur dilakukan dengan metode TF-IDF. Metode ini sudah terbukti handal dalam mengekstraksi fitur [21]. Metode ini menghitung frekuensi munculnya sebuah kata dalam dataset dan tingkat kepentingan kata tersebut.

Langkah selanjutnya adalah pembuatan model. Model dibuat dengan data twitter yang telah diolah pada langkah

sebelumnya. Data twitter akan dibagi menjadi data tes dan data uji. Data tes bitentukan sebesar 70% dari total data. Sisa data sebanyak 30% akan digunakan sebagai data uji.

Akan ada beberapa model yang dibuat. Tiap model dibuat dengan metode yang sudah umum digunakan dalam analisis sentimen. Total akan ada empat metode yang akan dibuat modelnya. Metode yang akan digunakan sebagai model adalah SVM, Naïve Bayes, dan Logistic Regression. Metode-metode ini sudah terbukti handal pada penelitian sebelumnya.

Selain tiga metode tersebut, akan dibuat juga model dengan metode Random Forrest. Metode ini memiliki performa dan akurasi yang baik dalam mengklasifikasi sentiment [22]. Keempat metode akan dibuat modelnya dan dievaluasi mana yang terbaik. Hasil evaluasi akan dibahas pada sub bab Hasil dan Pembahasan.

Setelah model dibuat, model akan digunakan untuk melabeli data komentar facebook. Data ini telah mengalami *preprocessing* dan ekstraksi fitur dengan metode yang sama seperti data twitter. Akan ada dua label yang digunakan, positif dan negatif. Dari hasil pelabelan ini akan didapat sentimen masyarakat terhadap perkembangan kasus Covid 19 selama Januari dan Februari. Detail metodologi penelitian dapat dilihat pada Gambar 3.



Gambar 1. Metode penelitian secara umum

#### IV. HASIL DAN PEMBAHASAN

Hasil Data twitter yang digunakan adalah data yang sudah dilabeli dengan sentimen positif, negatif, dan netral. Karena model yang dibuat hanya mengklasifikasikan sentimen menjadi positif dan negatif, maka data untuk

sentimen netral dihapus. Sisanya adalah dataset yang berisi 5479 baris kalimat. Dataset tersebut terdiri dari 2887 sentimen negatif dan 2592 positif.

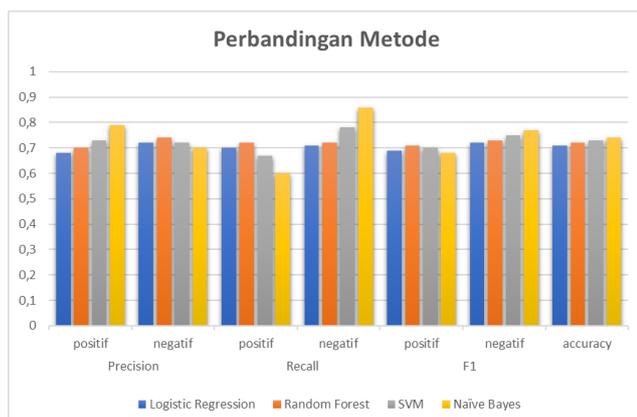
Data tersebut selanjutnya dilakukan *preprocessing* dan ekstraksi fitur menggunakan TF-IDF. Setelah itu, data dibagi dua menjadi data train dan data tes. Data train digunakan untuk pembuatan model dengan metode yang telah ditentukan. Model tersebut kemudian dievaluasi menggunakan data tes.

Model yang telah dibuat dievaluasi dengan *confusion matrix*. Cara ini dinilai cukup bagus dalam mengevaluasi suatu model [23]. Tiap model dicari nilai *confusion matrix* nya. Kemudian, dari *confusion matrix* yang didapat dicari nilai *precision*, *recall*, *accuracy*, dan nilai F1. *Recall* menunjukkan rasio jumlah dari suatu kelas dibandingkan dengan total jumlah elemen dari kelas tersebut. *Precision* menunjukkan rasio prediksi yang benar dari kelas tersebut. *Accuracy* menunjukkan keakuratan model dalam memprediksi. F1 menunjukkan rata-rata harmonik antara *recall* dan *precision*. Hasil evaluasi model dapat dilihat pada Tabel 1.

Tabel 1. Hasil evaluasi model

Metode	Sentimen	Precision	Recall	F1	Accuracy
Logistic Regression	Negatif	0.72	0.71	0.72	0.71
	Positif	0.68	0.70	0.69	
Random Forest	Negatif	0.74	0.72	0.73	0.72
	Positif	0.70	0.72	0.71	
SVM	Negatif	0.72	0.78	0.75	0.73
	Positif	0.73	0.67	0.70	
Naïve Bayes	Negatif	0.70	0.86	0.77	0.74
	Positif	0.79	0.60	0.68	

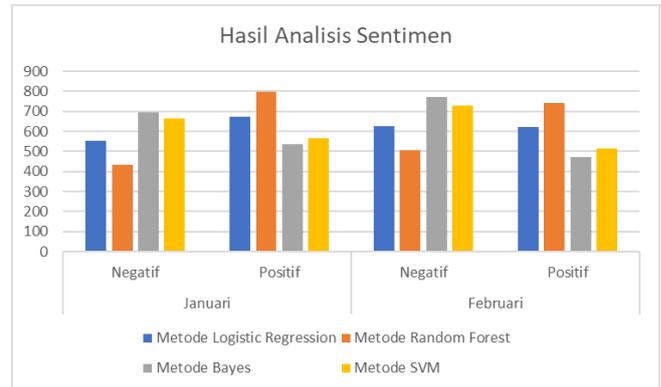
Dari Tabel 1 dapat disimpulkan bahwa Naïve Bayes adalah model dengan akurasi tertinggi. Namun, perbedaan F1 yang tinggi membuat Naïve Bayes lebih bagus untuk memprediksi sentimen negatif daripada positif. Hal ini juga berlaku pada SVM. Di sisi lain, walau nilai akurasi lebih rendah, Random Forest dan Logistic Regression memiliki nilai F1 yang lebih seimbang. Hal ini menunjukkan bahwa dua model tersebut lebih netral daripada dua model sebelumnya. Perbandingan performa metode dapat dilihat pada Gambar 4.



Gambar 2. Grafik perbandingan metode

Model-model yang telah dibuat kemudian dipakai untuk

mengklasifikasikan komentar facebook. Ada total 2476 komentar facebook yang diklasifikasi. Komentar tersebut terdiri dari 1229 komentar pada bulan Januari dan 1247 komentar pada bulan Februari. Hasil klasifikasi bisa dilihat pada Tabel 2 dan perbandingannya bisa dilihat pada Gambar 5.



Gambar 3. Hasil analisis sentimen

Dari Tabel 2 dapat dilihat bahwa pada bulan Januari, Naïve Bayes dan SVM lebih banyak mendeteksi sentimen negatif. Sebaliknya, Logistic Regression dan Random Forest lebih banyak mendeteksi sentimen positif. Selanjutnya pada bulan Februari, Naïve Bayes masih cenderung lebih banyak memprediksi sentimen negatif daripada positif. Sebaliknya, Random Forest cenderung memprediksi sentimen positif. Di sisi lain, Logistic Regression cenderung netral dalam memprediksi.

Tabel 2. Hasil analisis komentar Facebook

Bulan	Sentimen	Metode			
		Logistic Regression	Random Forest	Bayes	SVM
Januari	Negatif	555	432	694	664
	Positif	674	797	535	565
Februari	Negatif	626	507	774	731
	Positif	621	740	473	516

## V. KESIMPULAN

Pada penelitian ini telah dibuat beberapa model untuk memprediksi komentar facebook terhadap perkembangan COVID 19 selama bulan Januari dan Februari. Model dibuat dengan metode-metode yang sudah umum digunakan dalam penelitian sebelumnya. Selain itu, dibuat juga model dengan metode Random Forest yang jarang digunakan dalam memprediksi sentimen terkait COVID 19. Random Forest terlihat lebih netral dalam memprediksi sentimen.

Hasil prediksi sentimen komentar facebook dengan model-model yang dibuat menghasilkan klasifikasi sentimen dengan rasio antara negatif dan positif yang berbeda-beda. Walaupun begitu, rasio antara sentimen positif dan negatif tidak terlalu berbeda jauh. Hal ini menunjukkan bahwa masyarakat masih terbagi antara yang bersentimen positif terhadap perkembangan COVID 19 dengan yang negatif. Ini dapat menjadi tantangan bagi pemerintah untuk mengambil kebijakan yang seimbang dan memuaskan dua pihak.

Penelitian ini menggunakan data twitter untuk melatih model yang dibuat. Namun, data tersebut adalah data untuk sentimen secara umum. Performa model akan dapat

ditingkatkan lagi jika terdapat data yang khusus dibuat untuk mengklasifikasikan sentimen pada topik COVID 19. Akan lebih baik lagi jika data tersebut berupa data yang berasal dari komentar facebook. Hal ini dikarenakan facebook jauh lebih banyak digunakan oleh masyarakat Indonesia dibanding media sosial lainnya.

## VI. REFERENSI

- [1] Y.-C. Wu, C.-S. Chen, and Y.-J. Chan, "The outbreak of COVID-19: An overview," *Journal of the Chinese Medical Association*, vol. 83, no. 3, pp. 217–220, Mar. 2020, doi: 10.1097/JCMA.0000000000000270.
- [2] D. Susanna, "When will the COVID-19 Pandemic in Indonesia End?," *Kesmas: National Public Health Journal*, vol. 15, no. 4, Nov. 2020, doi: 10.21109/kesmas.v15i4.4361.
- [3] *Kesan Masyarakat Setelah Setahun Pandemi Covid-19*, (Mar. 02, 2021). Accessed: May 18, 2021. [Online Video]. Available: <https://www.cnnindonesia.com/tv/20210302161938-407-612757/video-kesan-masyarakat-setelah-setahun-pandemi-covid-19>
- [4] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS*, vol. 11, no. 2, p. 187, Jul. 2017, doi: 10.22146/ijccs.24716.
- [5] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," *JISEBI*, vol. 6, no. 2, p. 112, Oct. 2020, doi: 10.20473/jisebi.6.2.112-122.
- [6] Ridlo, Ilham, Putri, Nuzulul, Zein, Rizqy, Rahvy, Aisyah, and Junaidi, Erika, "Expression of Public Confidence in Indonesian Government in Handling COVID-19 Pandemic: Sentiment Analysis and Critical Content on Twitter," 2020, doi: 10.17605/OSF.IO/792NT.
- [7] 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," *jsi*, vol. 15, no. 1, pp. 27–29, Nov. 2020, doi: 10.30864/jsi.v15i1.332.
- [8] H. M. Denny, "Overview of new normal at work during COVID-19 pandemic in Indonesia: Social media and online news issues," vol. 24, no. 01, p. 10, 2021, doi: <http://doi.org/10.36295/ASRO.2021.24148>.
- [9] A. Pertiwi, A. Triayudi, and E. T. E. Handayani, "Sentiment Analysis of the Impact of Covid-19 on Indonesia's Economy through Social Media Using the ANN Method," vol. 4, no. 36, p. 8, 2020.
- [10] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *JIEET*, vol. 04, no. 02, p. 10, 2020, doi: <http://dx.doi.org/10.26740/jieet.v4n2.p90-92>.
- [11] S. Sautomo, N. Hafidz, and Y. E. Achyani, "SENTIMENT ANALYSIS DUE TO 'MUDIK' PROHIBITED OF COVID-19 THROUGH TWITTER," vol. 6, no. 1, p. 6, 2020.
- [12] [gs.statcounter.com](https://gs.statcounter.com), "Social Media Stats Indonesia." Accessed: May 18, 2021. [Online]. Available: <https://gs.statcounter.com/social-media-stats/all/indonesia>
- [13] T. B. T. Dewi, N. A. Indrawan, I. Budi, A. B. Santoso, and P. K. Putra, "Community Understanding of the Importance of Social Distancing Using Sentiment Analysis in Twitter," in *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*, Yogyakarta, Indonesia, Sep. 2020, pp. 336–341. doi: 10.1109/IC2IE50715.2020.9274589.
- [14] A. K. Fauziyyah, "ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON," *JIS*, vol. 18, no. 2, p. 31, Jul. 2020, doi: 10.30646/sinus.v18i2.491.
- [15] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regression," in *2020 6th Information Technology International Seminar (ITIS)*, Surabaya, Indonesia, Oct. 2020, pp. 238–242. doi: 10.1109/ITIS50118.2020.9320958.
- [16] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 8, no. 4, p. 334, Nov. 2019, doi: 10.22146/jnteti.v8i4.533.
- [17] J. H. Jaman, R. Abdulrohman, A. Suharso, N. Sulistiowati, and I. P. Dewi, "Sentiment Analysis on Utilizing Online Transportation of Indonesian Customers Using Tweets in the Normal Era and the Pandemic Covid-19 Era with Support Vector Machine," *Adv. sci. technol. eng. syst. j.*, vol. 5, no. 5, pp. 389–394, 2020, doi: 10.25046/aj050549.
- [18] S. Majumder, A. Aich, and S. Das, "Sentiment Analysis of People During Lockdown Period of COVID-19 Using SVM and Logistic Regression Analysis," *SSRN Journal*, 2021, doi: 10.2139/ssrn.3801039.
- [19] KawalCOVID.id, <https://www.facebook.com/KawalCOVID19>
- [20] A. Librian, "High quality stemmer library for Indonesian Language (Bahasa)," Feb. 21, 2017. <https://github.com/sastrawi/sastrawi>
- [21] A. I. Kadhim, "Term Weighting for Feature Extraction on Twitter: A Comparison Between BM25 and TF-IDF," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, Zakho - Duhok, Iraq, Apr. 2019, pp. 124–128. doi: 10.1109/ICOASE.2019.8723825.
- [22] A. Gupte, S. Joshi, P. Gadgul, A. Kadam, and A. Gupte, "Comparative study of classification algorithms used in sentiment analysis," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 5, pp. 6261–6264, 2014.
- [23] J. Lever, M. Krzywinski, and N. Altman, "Classification evaluation," *Nat Methods*, vol. 13, no. 8, pp. 603–604, Aug. 2016, doi: 10.1038/nmeth.3945.