

ANALISIS SENTIMENT COVID-19 DI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SVM

Sri Hadiani¹, Firman Yosep Tember²

¹Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

Jl. Raya Jatiwaringin No 2, Cipinang Melayu, Makasar, Jakarta Timur

²Program Studi Teknik Informatika, Fakultas Teknologi Informasi, ARS University

Jl. Sekolah Internasional, Antapani, Kiara Condong, Bandung

sri.shv@nusamandiri.ac.id, firmanyoseptember07@gmail.com

Abstract - People's aspirations are very difficult to convey directly, because it is constrained by several factors, such as time, space and other factors. However, social media is now used as a medium for voicing suggestions, criticisms and opinions to the general public, one of the social media used is Twitter. However, opinions on Twitter have a lot of different meanings from each user, one of which is about COVID-19. Therefore, the purpose of this study is to analyze public sentiment from Twitter tweets about COVID-19 in Indonesia, the methods used are Naive Bayes and Support Vector Machine. This study focuses on comparing the results of the classification of the two methods, as well as to determine the tendency of public opinion on Twitter. Subjects were taken using a public dataset with the keyword COVID-19 dataset. starting from the classification of positive and negative opinions, and get the final result. From the test results, it is known that the Support Vector Machine method has the highest level of accuracy with an accuracy rate of 54.21% and there is also a tendency for public opinion throughout the world to tend to be negative, this can be seen from the positive opinion of 98 and negative of 116. While the Naïve Bayes method obtained an Occurance rate of 53.27%.

Keywords - Covid19, Naïve Bayes, SVM, Sentiment, Twitter.

Abstrak - Aspirasi masyarakat sangat sulit disampaikan secara langsung, karena terkendala beberapa faktor, seperti waktu, ruang dan faktor lainnya. Namun, media sosial sekarang digunakan sebagai media menyuarakan saran, kritik dan pendapat kepada masyarakat umum, salah satu media sosial yang digunakan adalah twitter. Namun opini di twitter banyak sekali makna yang berbeda dari setiap pengguna, salah satu nya mengenai COVID-19. Oleh karena itu tujuan penelitian ini adalah menganalisis sentimen publik dari cuitan Twitter mengenai COVID-19 di Indonesia, metode yang digunakan yaitu Naïve Bayes dan Support Vektor Machine. Penelitian ini berfokus pada perbandingan hasil klasifikasi kedua metode tersebut, serta untuk mengetahui kecenderungan opini masyarakat di Twitter. Subjek diambil menggunakan dataset *public* dengan kata kunci dataset COVID-19. Dimulai dari klasifikasi opini positif dan negative, dan didapatkan hasil akhir. Dari hasil pengujian diketahui metode Support Vektor Machine memiliki tingkat akurasi yang paling tinggi dengan tingkat akurasi sebesar 54.21 % dan didapatkan pula kecenderungan opini masyarakat di seluruh duni condong negatif, hal tersebut dapat dilihat dari opini positif sebesar 98 dan negative sebesar 116. Sedangkan metode Naïve Bayes memperoleh tingkat Occurance sebesar 53.27%.

Kata Kunci - Covid 19, Naïve Bayes, SVM, Sentiment, Twitter.

I. PENDAHULUAN

Coronavirus 19 (COVID-19) merupakan infeksi dari gangguan pernapasan coronavirus 2 (SARS-CoV-2) yang dapat menular dengan cepat [1]. COVID-19 adalah salah satu penyakit yang bermula di daerah Wuhan, Tiongkok sejak Bulan November 2019. Sejak kejadian luar biasa muncul di Wuhan Cina, pada Desember 2019, World Health Organization (WHO) kemudian memberi nama coronavirus jenis baru yang ditemukan pada manusia tersebut sebagai Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) yang menjadi penyebab dari Coronavirus Disease-2019 (COVID-19) [2]. Penyebaran

COVID-19 sudah menyita perhatian dunia, terutama dunia kesehatan dimana dapat mengganggu

kesehatan fisik dan mental manusia [3]. Virus Covid-19 sudah masuk ke Indonesia dari awal maret sampai saat ini dan diperkirakan akan berlanjut sampai waktu yang belum diketahui [4].

Terdapat kurang lebih 200 negara yang telah terkena virus ini dari yang terbanyak yaitu Amerika Serikat dengan kurang lebih 186.046 orang positif corona, lalu Italia dengan 105.792 kasus, dan Spanyol sebanyak 95.923 kasus. Negara Indonesia sendiri pada 10 April 2020 terdapat 3.512 kasus yang positif, sembuh 282 orang dan meninggal sebanyak 306 orang dengan fatality rate atau tingkat kematian sebesar 9,1% [5]. Masyarakat menjadi takut dan khawatir dengan virus Covid-19, banyak masyarakat yang ingin meluapkan aspirasi terutama di jejaring sosial yang dirasa cocok dan tepat untuk mewakili aspirasi masyarakat, salah satu media sosial yang sering

digunakan adalah *twitter*, dimana *twitter* memungkinkan pengguna mengirim karakter hingga 140 karakter [6].

Komentar yang diberikan banyak sekali, ada komentar yang positif adapun komentar yang negatif sehingga salah satu cara yang dapat digunakan adalah analisis sentiment [7]. Analisis sentimen merupakan salah satu domain penting dalam machine learning [8]. Studi yang melakukan analisis tentang sentiment pendapat masyarakat di Twitter mengenai Covid-19 berdasarkan pada penelitian [6]. Penelitian tersebut menyebutkan bahwa hasil pengujian menggunakan metode Naïve Bayes opini masyarakat cenderung positif. Penelitian lainnya juga dilakukan [9], [10], [11]. Oleh karena itu, pada penelitian ini dilakukan analisis sentiment dengan metode klasifikasi Naïve Bayes Classification dengan ditambahkan metode SVM sebagai perbandingan metode. Tujuan penelitian ini adalah menganalisis sentimen publik dari cuitan Twitter mengenai COVID-19 di Indonesia. Penelitian ini diharapkan memberikan manfaat yaitu menambah pengetahuan mengenai analisis sentimen publik dari twitter sehingga dapat menjadi referensi pada penelitian selanjutnya yang berhubungan dengan topik terkait. Adapun keterbaruan penelitian ini dari penelitian sebelumnya adalah penggunaan keyword ng COVID-19. Data yang diambil dari website Kaggle sebanyak 228 data twit yang terdiri dari twit positif dan negative [12]. Data public tersebut sudah dikelompokkan antara positif, negative, dan netral dilihat dari twit pengguna di Indonesia. Cuitan/tweet tersebut dijadikan objek untuk melakukan analisis sentiment pada penelitian ini dengan menggunakan Teknik analisis klasifikasi menggunakan metode naïve bayes dan SVM.

Tabel 1. Tampilan data penelitian

Text	Sentiment
Yuppp Karena UTANG Pemerintah jg UTANG Bangsa INDONESIA Berarti jg HUTANG Nya AKU Sebagai Captain COVID 19 Kalo Orang Lain Itu TERSERAH Mereka..but AKU ya ini..apapun KEADAAN Pemerintah MEREKA TETAP BALA TENTARA ALLAH SWT Berarti SAHABAT Baik Nya Aku	Negative
Yuks.. kawal kebijakan pemerintah jangan sampai disalah gunakan oknum2 tidak	Positive

yang mengacu pada nama kebijakan pemerintah terkait penanganan COVID-19 serta dilihat dari segi waktu pengumpulan tweet yang lebih terkini.

II. METODE PENELITIAN

Tahap yang dilakukan dalam penelitian ini terdiri dari beberapa tahap:

1. Data

Data yang digunakan dalam penelitian ini terdiri dari data Twitter yang merupakan tweet berbahasa Indonesia hasil post dari wilayah Indonesia yang didapat dari website www.Kaggle.com. Website Kaggle merupakan website yang menyediakan kumpulan dataset dari berbagai bidang dengan tujuan untuk menunjang penelitian dari peneliti di berbagai belahan dunia. Tabel 1 merupakan beberapa tampilan data yang digunakan dalam penelitian.

2. Ruang Lingkup Penelitian

Mengambil data pada repositori Kaggle dengan menggunakan data twitter tenta

bertanggung jawab. Indonesia Menang Lawan Covid-19	
Yukk sama-sama bahu membahu membantuu pemerintah memutus rantaii covid-19 dan memutus rantai kebenciann. Supaya indonesia normal kembali	Positive

3. Teknik Analisis

Tahapan ini dilakukan klasifikasi (supervised learning) yang berupa teknik statistik dengan tujuan dapat mengelompokkan data pada kelas-kelas berlabel dengan tujuan membangun model berdasarkan pada data training serta melakukan prediksi kelas dari data yang baru [13]. Pengertian lainnya dari supervised learning (analisis klasifikasi) merupakan metode klasifikasi di mana kumpulan data sepenuhnya diberi label dimana selanjutnya dilakukan klasifikasi ke dalam kelas yang tidak dikenal [14]. Tujuan utama supervised learning yaitu mempelajari fungsi target yang dapat digunakan dalam memprediksi nilai dari suatu kelas [15].

Metode yang digunakan dalam Teknik klasifikasi adalah menggunakan metode Naïve Bayes dan SVM, dimana metode Naïve Bayes merupakan metode yang digunakan untuk mengelompokkan data sehingga dapat

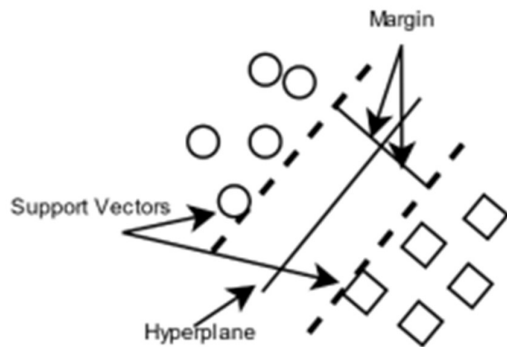
dilakukan prediksi probabilitas dari suatu class. Naïve Bayes merupakan metode pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh seorang ilmuwan inggris bernama Thomas Bayes [16]. Naïve Bayes memiliki argument dimana variabel prediktornya merupakan independen, artinya keberadaan fitur tertentu pada kelas tidak terkait dengan keberadaan fitur lainnya [17]. Adapun Teorema Bayes dapat dilihat pada rumus (1)

$$P(H|X) = \frac{P(H).P(H)}{P(X)} \dots\dots\dots(1)$$

Dimana:

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data merupakan suatu kelas spesifik
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)
- P(H) : Probabilitas hipotesis H (probabilitas prior)
- P(H|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas

Metode Support Vector Machine merupakan metode pembelajaran mesin yang termasuk ke dalam supervised learning. Support Vector Machine dapat berjalan dengan menentukan hyperplane paling baik dimana hasilnya memaksimalkan margin antara kelas yang berdekatan [18]. Pada Gambar 1 dapat dilihat ilustrasi dari metode Support Vector Machine.



Gambar 1. Ilustrasi SVM

4. Evaluasi dan Validasi

Setelah dilakukan kllasifikasi maka dilakukan evaluasi model yang meliputi akurasi, presisi, dan sensitivitas pada data *testing* dengan tujuan untuk mengukur ketepatan klasifikasi yang disebut dengan akurasi. Akurasi merupakan ukuran ketepatan klasifikai dimana dapat menunjukkan performansi teknik klasifikasi secara keseluruhan, dengan asumsi semakin tinggi akurasi klasifikasi semakin baik performansi teknik klasifikasi. Adapun tabel *confusion matrix* sebagai dasar

perhitungan evaluasi model dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	Tp	Fn
Negatif	Fp	Tn

$$\text{Akurasi Klasifikasi (\%)} = \frac{Tp+Tn}{Tp+Fp+Tn+Fn}$$

$$\text{Sensitifity (\%)} = \frac{Tp}{Tp+Fn}$$

$$\text{Specificity (\%)} = \frac{Tn}{Fp+Tn}$$

5. Visualisasi

Visualiasasi pada penelitian ini menggunakan *Wordcloud*. *Wordcloud* menampilkan kata yang menjelaskan setiap kelompok klasifikasi.

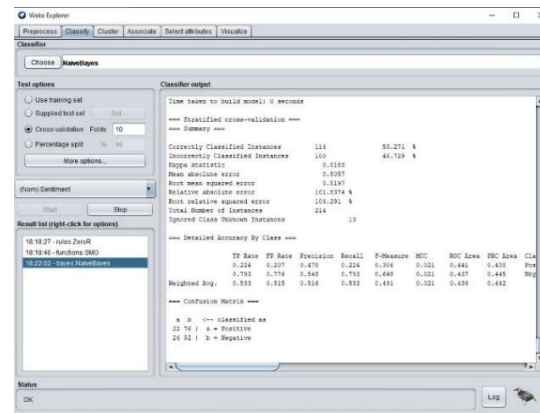
III. HASIL DAN PEMBAHASAN

a. Pembentukan Data *Testing* dan *Training*

Pada penelitian ini membagi data menjadi dua yaitu data training dan data testing dengan pembagian 80:20. Berdasarkan pembagian tersebut, diperoleh sebanyak 182 tweet yang digunakan sebagai data training dan sebanyak 46 tweet sisanya digunakan sebagai data testing. Setelah pembagian data menjadi dua telah dilakukan,kemudian dilakukan evaluasi model.

b. Klasifikasi

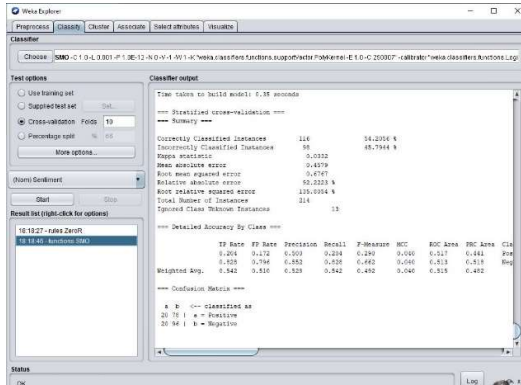
Gambar 2 dapat dilihat hasil dari pengujain data menggunakan naïve bayes



Gambar 2. Klasifikasi Naïve Bayes

Gambar 3 dapat dilihat hasil dari pengujian data menggunakan SVM

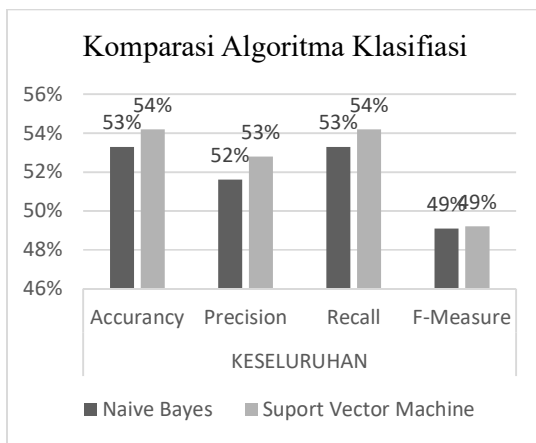
Setelah data semua data diperoleh, selanjutnya dilakukan dengan mengolah data menggunakan aplikasi WEKA versi 3.9.4 data diolah dan diperoleh hasil dari masing-masing model klasifikasi secara detail seperti Gambar 4:



Gambar 3. Klasifikasi SVM

	DETAIL PER OPINI					
	Accuracy		Precision		Recall	
	Positif	Negatif	Positif	Negatif	Positif	Negatif
Naive Bayes	53.20%	46.73%	48%	55%	22%	79%
Support Vector Machine	54.21%	45.79%	50%	55%	20%	83%
	KESELURUHAN					
	Accuracy	Precision	Recall	F-Measure		
Naive Bayes	53%	52%	53%	49%		
Support Vector Machine	54%	53%	54%	49%		

Gambar 4. Hasil Klasifikasi



Gambar 5. Komparasi Algoritma Klasifikasi

Gambar 5 dapat dilihat jika hasil klasifikasi model Naive Bayes lebih baik dibandingkan model Naive Bayes, ditinjau dari:

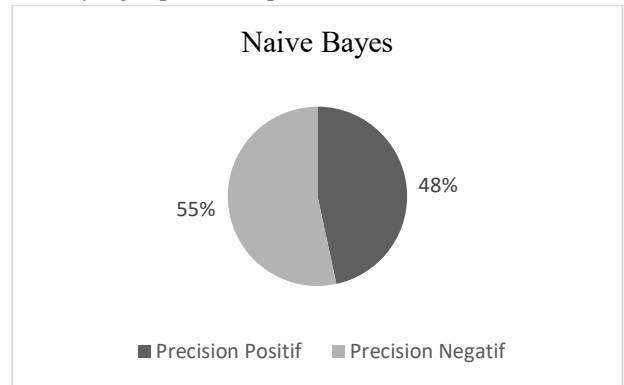
1. Nilai accuracy yang didapatkan dari pengujian Naive Bayes sebesar 53% sedangkan SVM sebesar 54%.

2. Nilai precision yang didapatkan dari metode Naive Bayes dengan nilai 52% dan SVM 53%.
3. Nilai recall dari data diatas didapatkan Naive Bayes 53% dan SVM 54%.
4. Nilai F-Measure didapatkan dari Naive Bayes 49% dan SVM 49%.

Sehingga dapat disimpulkan bahwa model klasifikasi SVM memiliki tingkat akurasi tinggi dari semua pengujian, dalam mengklasifikasikan data, dibandingkan dengan model klasifikasi Naive Bayes.

c. Visualisasi

Untuk mengetahui kecenderungan opini yang diuji bersifat positif atau negatif, dapat dilihat pada banyaknya opini yang ada, yaitu positif 98, dan opini negatif sebanyak 116, untuk memperkuat data kecenderungan opini yang diuji dapat dilihat pada hasil pengujian model SVM, pada bagian *precision*, karena menurut Indrayuni *precision*, dapat memprediksi ketepatan data dari suatu sistem. Diketahui *precision* negatif lebih banyak persentasenya dibanding *precision* positif, yaitu 55% : 50%, yang dapat dilihat pada Gambar 7.



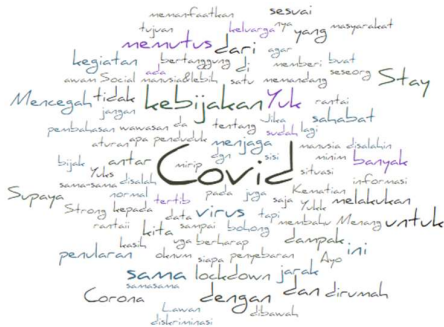
Gambar 7. Kecenderungan negative data opini yang diuji

Sehingga opini atau *tweets* masyarakat di twitter tentang COVID-19 saat ini, cenderung pada opini negative. Melihat kata yang sering muncul pada sentiment positif, dan negative dapat dilihat sebagai berikut:

1) *Wordcloud*. Sentimen Positif

Kata-kata sentimen positif yang disajikan dalam bentuk wordcloud dapat dilihat pada Gambar 8. Kata-kata yang menunjukkan bentuk pernyataan positif pengguna twitter terhadap COVID-19. Ukuran kata pada wordcloud menunjukkan banyaknya frekuensi kata yang muncul, dimana semakin besar ukuran kata, semakin tinggi frekuensi kata tersebut digunakan oleh pengguna twitter.

Berdasarkan wordcloud sentimen positif, kata yang paling sering muncul adalah kata “covid”, kemudian diikuti dengan kata “kebijakan”, “memutus” dan “stay”. Hal tersebut diakibatkan pada masa tersebut covid sedang tinggi-tingginya dan kebijakan pemerintah yang membuat masyarakat harus berdiam diri di rumah dan dibatasi aktifitasnya, sehingga kata “covid”, ‘kebijakan’, ‘memutus’, dan stay menjadi kata-kata yang bernada positif.



Gambar 8. Wordcloud. Sentimen Positif

2) Wordcloud. Sentimen Negatif

Kata-kata sentimen negatif yang disajikan dalam bentuk wordcloud dapat dilihat pada Gambar 9. Kata-kata yang menunjukkan bentuk pernyataan negatif pengguna twitter terhadap COVID-19. Ukuran kata pada wordcloud menunjukkan banyaknya frekuensi kata yang muncul, dimana semakin besar ukuran kata, semakin tinggi frekuensi kata tersebut digunakan oleh pengguna twitter. Berdasarkan wordcloud sentimen negatif, kata yang paling sering muncul adalah kata “pemerintah”, kemudian diikuti dengan kata “hutang” dan “covid”. Hal tersebut diakibatkan pada masa tersebut banyak masyarakat yang kecewa terhadap pemerintah, dan banyaknya isu bahwa membengkaknya hutang di negara Indonesia yang diakibatkan oleh Covid, sehingga banyak masyarakat yang membuat twit “pemerintah”, “hutang”, dan “covid” bernada negative.



Gambar 9. Wordcloud. Sentimen Negatif

IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat ditarik kesimpulan bahwa dari metode yang diujikan tersebut metode Suppor Vektor Machine memiliki tingkat accuracy yang paling tinggi dengan tingkat accuracy sebesar 54.21 % dan didapatkan pula kecenderungan opini masyarakat di seluruh duni condong negatif,hal tersebut dapat dilihat dari opini positif sebesar 98 dan negative sebesar 116. Sedangkan metode Naïve Bayes memperoleh tingkat accuracy sebesar 53.27%.

DAFTAR PUSTAKA

[1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses,” *Journal of Advanced Research*, vol. 24. Elsevier B.V., pp. 91–98, Jul. 01, 2020, doi: 10.1016/j.jare.2020.03.005.

[2] L. Yang *et al.*, “Epidemiological And Clinical Features Of 200 Hospitalized Patients With Corona Virus Disease 2019 Outside Wuhan, China: A Descriptive Study,” *J. Clin. Virol.*, vol. 129, Aug. 2020, doi: 10.1016/j.jev.2020.104475.

[3] K. Handayani, E. H. Juningsih, D. Riana, S. Hadiani, A. Rifai, and R. K. Serli, “Measuring the Quality of Website Services covid19.kalbarprov.go.id Using the Webqual 4.0 Method,” in *Journal of Physics: Conference Series*, Nov. 2020, vol. 1641, no. 1, doi: 10.1088/1742-6596/1641/1/012049.

[4] M. P. Lestari, D. J. Winarsyah, and F. Hamami, “Peramalan Pertambahan Pasien Covid-19 Menggunakan Support Vector Regression Forecasting Growth Of Covid-19 Patients Using Support Vector Regression,” 2021.

[5] E. Kartika Sari, B. Ria EMD, M. Karina Putri, M. Eka Rosita, P. S. Studi, and S. Tinggi Ilmu Kesehatan Akbidyo, “Pangan Fungsional Sebagai Alternatif Penunjang Imun Di Masa Pandemi,” 2021.

- [6] E. T. Handayani and A. Sulistiyawati, "Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 32–37, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>.
- [7] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [8] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques," *J. Comput. Sci.*, vol. 14, no. 5, pp. 714–726, 2018, doi: 10.3844/jcsp.2018.714.726.
- [9] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn," *Inti Nusa Mandiri*, 2020, doi: 10.33480/inti.v15i1.1347.
- [10] Samsir *et al.*, "Naives Bayes Algorithm for Twitter Sentiment Analysis," *J. Phys. Conf. Ser.*, vol. 1933, no. 1, 2021, doi: 10.1088/1742-6596/1933/1/012019.
- [11] S. H. Sahir, R. S. Ayu Ramadhana, M. F. Romadhon Marpaung, S. R. Munthe, and R. Watrianthos, "Online learning sentiment analysis during the covid-19 Indonesia pandemic using twitter data," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1156, no. 1, p. 012011, 2021, doi: 10.1088/1757-899x/1156/1/012011.
- [12] 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," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, p. 112, 2020, doi: 10.20473/jjisebi.6.2.112-122.
- [13] A. Andreyestha and A. Subekti, "Analisa Sentiment Pada Ulasan Film Dengan Optimasi Ensemble Learning," *J. Inform.*, vol. 7, no. 1, pp. 15–23, 2020, doi: 10.31311/ji.v7i1.6171.
- [14] A. Roihan, P. Abas Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2019.
- [15] Y. C. Liang, Y. Maimury, A. H. L. Chen, and J. R. C. Juarez, "Machine learning-based prediction of air quality," *Appl. Sci.*, vol. 10, no. 24, pp. 1–17, 2020, doi: 10.3390/app10249151.
- [16] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naive bayes," *Inf.*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- [17] Y. Watequlis Syaifudin and D. Puspitasari, "Twitter Data Mining for Sentiment Analysis on Peoples Feedback Against Government Public Policy," *MATTER Int. J. Sci. Technol.*, vol. 3, no. 1, pp. 110–122, 2017, doi: 10.20319/mijst.2017.31.110122.
- [18] F. Taufiqurrahman, S. Al Faraby, and M. D. Purbolaksono, "Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10650–10659, 2021.