

Analisis Sentimen Opini Masyarakat Indonesia Terhadap Vaksin Covid-19 Pada Sosial Media Twitter Menggunakan Metode Naïve Bayes dan k-Nearest Neighbors

Mitsal Shafiq Sulasno
Departemen Teknologi Informasi
Universitas Indonesia
Depok, Indonesia
mitsal.shafiq01@ui.ac.id

Hilda Amalia
Departemen Teknologi Informasi
Institut Perbanas
Jakarta, Indonesia
hilda21amalina@gmail.com

Valentinus Paramarta
Departemen Teknologi Informasi
Institut Perbanas
Jakarta, Indonesia
valentinus13@perbanas.id

Abstrak— Twitter merupakan salah satu media sosial yang digunakan untuk menuliskan opini mengenai vaksinasi Covid-19 di Indonesia. Penelitian ini melakukan analisis sentimen untuk mengetahui pandangan masyarakat Indonesia mengenai Vaksin Covid-19 apakah positif atau negatif. Penelitian ini menggunakan beberapa algoritma machine learning yang akan diuji performanya seperti algoritma Naïve Bayes dan k-Nearest Neighbors (kNN) untuk memprediksi sentimen masyarakat mengenai program vaksinasi Covid-19 melalui sosial media Twitter. Dari total 478 data set yang digunakan dalam penelitian ini, sebanyak 326 tweets merupakan respon positif dan 152 tweets respon negatif terhadap program vaksinasi Covid-19 di Indonesia. Berdasarkan hasil yang didapatkan, diketahui bahwa algoritma Naïve Bayes menunjukkan performa akurasi yang lebih baik dibandingkan dengan algoritma kNN dengan nilai akurasi yang didapatkan adalah sebesar 90%.

Kata Kunci—*machine learning, analisis sentimen, twitter, vaksinasi*

I. PENDAHULUAN

Saat ini seluruh dunia sedang dihadapi oleh pandemi *Coronavirus Disease* atau yang biasa disebut dengan Covid-19. Pada tanggal 11 maret 2020 WHO (*World Health Organization*) mendeklarasikan bahwa Covid-19 sebagai pandemi global. Pada tanggal 13 januari 2021, Indonesia menjadwalkan vaksinasi Covid-19 pertama dengan menggunakan vaksin Sinovac atau dikenal dengan

nama CoronaVac. Vaksinasi akan dijadwalkan untuk seluruh masyarakat Indonesia dengan dua tahap, tahap pertama dilakukan dalam rentang waktu Januari-April 2021 untuk para tenaga medis. Selanjutnya, mulai April 2021 hingga Maret 2022 dilaksanakan tahap kedua yang diperuntukkan bagi 63,9 juta masyarakat rentan dan 77,4 juta masyarakat lainnya yang diberikan sesuai pendekatan klaster.

Menurut laporan yang dipublikasikan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2019-2020, dari total 266.91 juta jiwa penduduk Indonesia terdapat 96.71 juta jiwa merupakan pengguna internet atau sekitar 73.7% dari total penduduk Indonesia. Twitter menjadi aplikasi ke-3 yang paling banyak digunakan oleh masyarakat Indonesia. Berdasarkan keseluruhan pengguna internet di Indonesia, diketahui bahwa sebanyak 10% penduduk Indonesia merupakan pengguna yang sering menggunakan Twitter dan 10.6% sesekali menggunakan Twitter [1]. Sama seperti pesaingnya (Facebook dan Instagram), Twitter sering digunakan oleh masyarakat Indonesia sebagai media pertukaran informasi dan pendapat termasuk mengenai vaksinasi.

Mengingat Twitter menjadi peringkat ke-3 sosial media yang paling sering digunakan oleh masyarakat Indonesia dalam periode 2019-2020, peneliti memanfaatkan peluang ini untuk mencari informasi dan menganalisa opini yang ada pada tweets masyarakat Indonesia mengenai program vaksinasi. Program vaksinasi di Indonesia ini menimbulkan reaksi dan opini

yang beragam dari masyarakat Indonesia dalam penggunaan sosial media Twitter, mulai dari yang setuju sampai yang tidak setuju dengan program pemerintah tersebut. Kelompok masyarakat yang setuju menganggap bahwa program vaksinasi merupakan hal penting karena diharapkan bisa melindungi diri dari virus Corona yang sedang menyebar, sedangkan masyarakat yang tidak setuju memiliki berbagai macam pendapat mulai dari kekhawatiran terhadap keamanan dan keefektifan vaksin hingga mempersoalkan kehalalan dari vaksin yang akan digunakan.

Salah satu teknologi yang dapat digunakan dan sedang berkembang saat ini untuk menganalisis opini masyarakat dalam penggunaan media sosial adalah teknologi *Machine Learning* dengan menggunakan teknik analisis sentimen. Analisis sentimen adalah teknik yang dilakukan untuk mendeteksi polaritas dari suatu teks yang bertujuan untuk menentukan apakah suatu teks yang dianalisis mengandung makna positif atau negatif. Penelitian terkait analisis sentimen vaksin Covid-19 juga telah dilakukan oleh beberapa peneliti terdahulu, seperti yang dilakukan oleh Syarifuddin (2020) dan Afshoh (2020) [2,3]. Namun, berdasarkan penelitian tersebut masih terdapat keterbatasan dalam tingkat akurasi dan algoritma *machine learning* yang digunakan. Sehingga pada penelitian ini, akan dilakukan penelitian terkait analisis sentimen masyarakat Indonesia terhadap program vaksinasi Covid-19 pada sosial media Twitter dengan menggunakan beberapa algoritma *machine learning* seperti Naïve Bayes dan *k-Nearest Neighbors* (kNN) untuk mengetahui performa terbaik dalam kasus yang akan diteliti.

Berdasarkan permasalahan yang telah dijelaskan di atas, maka rumusan permasalahan pada penelitian ini adalah:

RQ: Bagaimana perbandingan algoritma *machine learning* Naïve Bayes dan kNN yang tepat digunakan untuk menganalisis sentimen masyarakat Indonesia terhadap program vaksinasi Covid-19 melalui sosial media Twitter?

II. LANDASAN TEORI

Pada bagian ini akan dijelaskan terkait penelitian terdahulu yang telah dilakukan. Penelitian yang dijelaskan pada bagian ini adalah penelitian yang memiliki topik terkait penggunaan berbagai algoritma *machine learning* dalam melakukan tugas analisis sentimen di sosial media.

A. Penelitian Terdahulu Terkait Penggunaan Machine Learning Dalam Analisis Sentimen

Syarifuddin (2020) telah melakukan penelitian terkait analisis sentimen opini publik mengenai Covid-19 pada sosial media Twitter. Pada penelitian tersebut, dilakukan

perbandingan dua metode *machine learning* yaitu metode Naïve Bayes dan *k-Nearest Neighbors* (kNN). Berdasarkan penelitian tersebut didapatkan bahwa metode Naïve Bayes dapat memberikan performa tingkat akurasi yang lebih tinggi dibandingkan metode kNN yaitu sebesar 63.21%. Penelitian serupa juga dilakukan oleh Afshoh (2020) terkait penggunaan metode *machine learning* Naïve Bayes untuk melihat respon masyarakat terhadap kenaikan harga jual rokok. Pada penelitian ini digunakan dua metode yang berbeda yaitu metode Naïve Bayes dan Lexicon Based. Didapatkan pada penelitian tersebut bahwa metode Naïve Bayes memberikan performa akurasi yang lebih tinggi dibandingkan dengan metode Lexicon Based, yaitu masing-masing sebesar 80% dan 65% [2].

Penelitian yang serupa juga telah dilakukan oleh Kurniawan (2020), Garcia (2021), dan Raghupathi (2020). Dimana pada berbagai penelitian tersebut telah dilakukan penelitian terkait penggunaan *machine learning* dalam analisis sentimen masyarakat terhadap kasus Covid-19 dan vaksinasi Covid-19 melalui sosial media Twitter. Dalam berbagai penelitian tersebut didapatkan bahwa algoritma *machine learning* berhasil memberikan hasil yang baik dan sesuai dengan tujuan penelitian terkait analisis sentimen masyarakat [4,5,6].

B. Algoritma Machine Learning

Pada penelitian ini, algoritma *machine learning* yang akan digunakan adalah Naïve Bayes dan *k-Nearest Neighbors* (kNN). Pemilihan algoritma *machine learning* tersebut didasarkan oleh berbagai penelitian terdahulu yang telah dilakukan. Seperti yang dilakukan oleh Villavicencio (2021), dimana pada penelitian tersebut dilakukan penelitian analisis sentimen pada sosial media Twitter terkait vaksinasi Covid-19 di Philippines menggunakan algoritma Naïve Bayes. Pada penelitian tersebut didapatkan bahwa algoritma Naïve Bayes dapat digunakan untuk memprediksi sentimen masyarakat Philippines yang ditunjukkan dengan performa akurasi yang cukup baik yaitu 81.77% [7]. Penelitian yang serupa dengan menggunakan algoritma Naïve Bayes juga telah dilakukan oleh Ritonga (2021), Gautam (2021), dan Muzaki (2021) [8,9,10].

Algoritma *k-Nearest Neighbors* (kNN) juga merupakan salah satu algoritma *machine learning* yang sering digunakan dalam analisis sentimen. Seperti penelitian yang dilakukan oleh Isnain (2021) terkait analisis sentimen masyarakat terkait pembelajaran daring melalui sosial media Twitter. Dimana didapatkan dalam penelitian tersebut bahwa algoritma kNN dapat memberikan performa akurasi yang cukup baik yaitu sebesar 85% [11]. Penelitian yang serupa juga didapatkan pada penelitian lainnya seperti penelitian yang dilakukan

oleh Shamrat (2021), Kirana (2021), dan Masturoh (2021) [12,13,14].

III. METODOLOGI

Pada bagian ini akan dijelaskan data yang digunakan dalam penelitian, metode dan alat bantu yang digunakan, dan tahapan penelitian yang dilakukan.

A. Data Penelitian

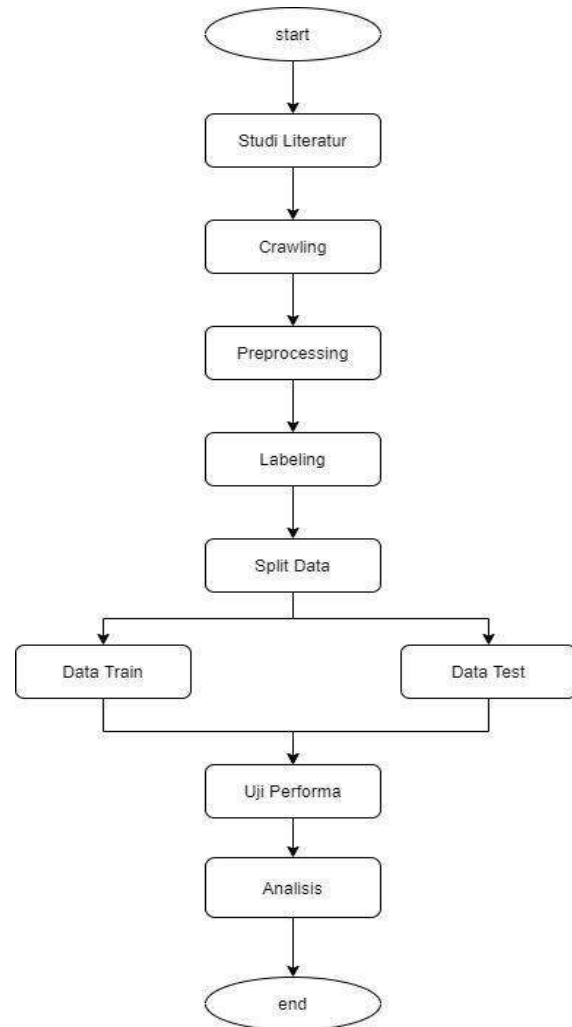
Sumber data pada penelitian ini adalah tweet para pengguna twitter indonesia yang membahas tentang vaksinasi Covid-19 di Indonesia. Peneliti mengambil sampel sebanyak 500 tweet dan hanya yang membahas vaksinasi Covid-19 di Indonesia. Periode tweet yang digunakan dalam penelitian ini adalah periode program vaksinasi tahapan pertama dan kedua dilaksanakan di Indonesia, yaitu mulai dari tanggal 1 Januari 2021 sampai dengan 1 Agustus 2021. Pemilihan pengambilan periode tweet tersebut dimaksudkan agar peneliti dapat menangkap opini masyarakat Indonesia yang bervariasi (pro dan kontra) karna program vaksinasi baru dimulai. Adapun fitur data yang digunakan dalam penelitian ini berasal dari kamus kata yang didapatkan dari tweet yang diambil.

B. Tahapan Penelitian

Penelitian ini dilakukan dengan melaksanakan beberapa tahapan penelitian dimulai dari studi literatur, penarikan data melalui API sosial media Twitter, *pre-processing data* (pembersihan data, mengubah kata tidak baku menjadi kata baku, menghapus karakter spesial ataupun simbol, transformasi data menjadi frekuensi kata yang muncul), pebelan data yang akan digunakan, pemecahan data menjadi data pelatihan dan data pengujian, uji performa algoritma *machine learning*, dan analisis hasil yang didapatkan seperti yang ditunjukkan pada Gambar 1.

a) *Studi Literatur*: Pada tahap ini dilakukan studi pustaka yang berhubungan dengan penelitian ini, seperti analisis sentimen, metode *machine learning* seperti Naive bayes dan kNN, serta literatur yang membahas topik yang sama dan digunakan peneliti sebagai tumpuan untuk penelitian ini.

b) *Crawling (Penarikan Data)*: Pada tahap ini, peneliti mengumpulkan data dari media sosial Twitter menggunakan API twitter dan mengumpulkan data menggunakan teknik *crawling* menggunakan bahasa pemrograman Python. Data ini merupakan tweets dari pengguna Twitter Indonesia yang membahas vaksinasi Covid-19. Kriteria data yang digunakan pada tahapan ini adalah sebagai berikut: (1) *tweet* yang di *posting* melalui sosial media Twitter mulai dari tanggal 1 Januari 2021 sampai dengan 1 Agustus 2021; (2) kolom data yang ditarik pada tahapan ini adalah kolom *id* (akun pengguna



Gambar 1. Tahapan Penelitian

Twitter), *created_at* (tanggal tweet di posting oleh pengguna), *retweet_count* (jumlah retweet), *hashtag* (*hashtag* yang digunakan pada *tweet* yang di *posting*), dan *place* (lokasi *tweet* di *posting*); (3) kata kunci yang digunakan adalah ‘vaksinasi’, sehingga setiap *tweet* yang disebar pada rentang waktu yang telah ditentukan serta terdapat kata kunci yang digunakan maka *tweet* tersebut akan dijadikan sebagai sumber data penelitian ini; (4) *tweet* yang digunakan sebagai sumber data penelitian adalah *tweet* yang menggunakan Bahasa Indonesia.

c) *Pre-processing*: Setelah melewati tahapan penarikan data (*crawling data*), data perlu dibersihkan terlebih dahulu sebelum data diolah lebih lanjut dalam tahapan selanjutnya. Terdapat beberapa tahapan *pre-processing data* yang dilakukan dalam penelitian ini. Pertama, dilakukan penghapusan terkait kata pada *tweet* yang mengandung kata kunci yang digunakan pada

tahapan *crawling data*, yaitu adalah kata ‘vaksinasi’ atau ‘vaksin’. Penghapusan kata kunci yang digunakan pada *tweet* yang dikumpulkan dilakukan dengan tujuan supaya *tweet* yang akan digunakan hanya mengandung kosa kata yang bermakna positif atau negatif saja. Sedangkan kata kunci ‘vaksin’ atau ‘vaksinasi’ tidak mengandung makna positif ataupun negatif sehingga dapat dihilangkan dari set data *tweet*.

Kedua, merupakan tahapan tokenisasi dimana tahapan ini dilakukan untuk mendapatkan semua kata yang ada pada *tweet* yang dikumpulkan pada tahapan sebelumnya. Kumpulan kata tersebut selanjutnya akan dikumpulkan dalam kumpulan kata atau kamus kata. Ketiga, dilakukan proses normalisasi untuk mengubah kata tidak baku menjadi kata baku yang terdapat pada kamus kata yang telah dihasilkan dari proses sebelumnya. Selain itu, pada proses normalisasi ini juga akan dilakukan transformasi huruf menjadi huruf kecil secara keseluruhan serta menghapus tanda baca yang ada seperti titik, koma, spasi, dan sebagainya.

Keempat, merupakan proses *stop-word removal* atau menghapus kata *stop-word*. Proses ini dilakukan untuk menghapus kata-kata pada kamus kata yang memiliki frekuensi muncul paling banyak dan paling sedikit untuk mengurangi kata-kata yang tidak memiliki makna signifikan pada *tweet* yang dijadikan data set. Kelima merupakan proses *stemming*. Dimana proses ini dilakukan untuk mengubah kata-kata yang memiliki makna yang sama menjadi suatu kata yang dapat dianggap mewakili kata tersebut.

d) Labeling

Proses *labeling* dilakukan setelah seluruh *tweets* yang diambil selesai melewati tahap *preprocessing*. Pada proses ini dilakukan *labeling* terhadap kata dari *tweets* apakah mengandung makna positif atau negatif. Proses *labeling* dilakukan secara otomatis berdasarkan *lexicon* (*dictionary*) yang berjumlah 10.287 kata. Dimana dalam kamus *lexicon* yang digunakan dalam penelitian ini terdapat daftar kata yang mengandung kata positif maupun negatif serta bobot kata tersebut. Selanjutnya, kamus *lexicon* akan disesuaikan dengan kamus kata yang digunakan dari proses *preprocessing* sebelumnya sehingga akan dihasilkan label untuk setiap *tweet* dalam data set.

Tahap selanjutnya adalah menghitung bobot dari tiap *term* pada setiap kelas sentimen dengan menggunakan *Term Frequency – Inverse Document Frequency* (TFIDF). *Weight Term Document* atau biasa yang disebut TF-IDF dari suatu *term* atau token atau kata merupakan hasil perkalian antara *tf weight* dengan *idf*.

e) Split Data

Setelah dilakukan pembobotan *term* dengan menggunakan *Term Frequency – Inverse Document Frequency* (TFIDF), selanjutnya data set akan dipecah menjadi data pelatihan dan data pengujian sebelum dilakukan evaluasi. Pembagian data pelatihan dan data pengujian yang digunakan dalam penelitian ini adalah masing-masing sebesar 80% dan 20%.

f) Uji Performa

Sebelum dilakukan pengujian performa terhadap algoritma *machine learning* yang akan digunakan, dilakukan proses validasi terlebih dahulu dengan menggunakan 10-fold cross validation. Dimana data akan terbagi secara acak menjadi 10 bagian data dengan jumlah yang sama dan akan dilakukan proses validasi sebanyak 10 kali secara berulang.

Selanjutnya akan dilakukan pengujian performa untuk mengetahui tingkat kebenaran proses klasifikasi yang ditabulasikan dalam suatu Tabel yang disebut dengan *confusion matrix*. Algoritma *machine learning* yang akan diuji dalam penelitian ini adalah Naïve Bayes dan k-Nearest Neighbors (kNN). Nilai yang akan dievaluasi dalam penelitian ini adalah nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

IV. HASIL DAN PEMBAHASAN

A. Set Data

Pengumpulan data adalah langkah kedua dalam penelitian ini, langkah ini dilakukan setelah melakukan studi literatur. Data yang menjadi bahan dalam penelitian ini diambil menggunakan Twitter API, pengambilan data ini disebut dengan *crawling*. Pada Tabel 1 adalah hasil dari proses *crawling*.

Tabel 1. Contoh Hasil Proses *Crawling*

No	<i>Tweet</i>
1	Vaksin Pertama Pelajar Oleh BIN Disambut Baik, Target 3000 Vaksin Diberikan Serentak
2	Ayo dukung Vaksin Nusantara dengan mengisi survey berikut ini. https://t.co/FURRePOuV0
3	belum vaksin 🤔 aku harus apa 🤔
4	WAH KEREN, SEHAT TERUS BUAT PERI DAN KELUARGA YAA. Kalau aku baru vaksin pertama 🤔👍
5	Dalam rangka mensukseskan akselerasi vaksinasi, Aliansi Badan Eksekutif Mahasiswa (BEM) kunjungi Mapolda Banten
6	abis di vaksin jdi pegel 🤔👍
7	klo emang ga percaya sama c0v1d yaudah, tp tetep vaksin n patuhin prokes dong 🤔👍

Tabel 1 diatas merupakan contoh sekumpulan *tweets* yang didapat dari proses *crawling*. *Tweets* diatas juga murni dari Twitter dan belum melewati proses *pre-processing* dimana dapat dilihat bahwa masih terdapat beberapa kata yang belum dalam bentuk baku, terdapat emoticon dalam *tweet*, terdapat penggunaan huruf besar dan huruf kecil, link, dan kata yang merupakan bentuk singkatan. Sehingga diperlukan untuk dilakukan pemrosesan lebih lanjut data mentah tersebut dalam tahapan *pre-processing*.

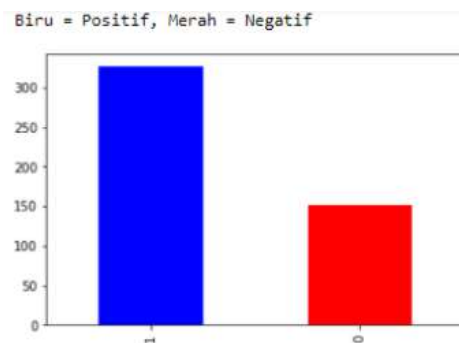
Berikut pada Tabel 2 merupakan contoh hasil data set setelah dilakukan tahapan *pre-processing* seperti perubahan huruf besar menjadi huruf kecil, perubahan kata tidak baku menjadi kata baku, penghapusan emoticon dalam *tweet*, dan penghapusan kata *stop-word* pada *tweet*.

Tabel 2. Contoh Hasil Tahapan *Pre-Processing*

No	Tweet
1	vaksin pertama pelajar oleh bin disambut baik target 3000 vaksin diberikan serentak
2	dukung vaksin nusantara dengan mengisi survey
3	belum vaksin aku harus apa
4	keren, sehat terus buat peri dan keluarga ya kalau aku baru vaksin pertama
5	dalam rangka sukseskan akselerasi vaksinasi aliansi badan eksekutif mahasiswa mengunjungi mapolda banten
6	habis di vaksin jadi pegal
7	kalau tidak percaya dengan covid tetapi tetap vaksin dan patuhin protokol kesehatan

Setelah melewati tahapan *pre-processing*, selanjutnya data set akan diberikan label untuk setiap *tweet* yang telah dikumpulkan. Tahapan pemberian label ini menggunakan kamus *lexicon* yang berikan kamus kata positif dan kamus kata negatif beserta dengan bobotnya. Setiap kata akan dihitung bobotnya berdasarkan kamus *lexicon* yang digunakan, hasil dari penghitungan bobot tersebut akan menjadi label untuk setiap *tweet* yang digunakan dalam data set.

Berdasarkan hasil penyaringan data set berdasarkan tahapan *pre-processing*, didapatkan bahwa *tweet* yang akan diproses lebih lanjut pada tahapan selanjutnya sebanyak 478 *tweet*. Hal tersebut dikarenakan terdapat beberapa *tweet* yang masih merupakan *re-tweet* dan tidak mengandung makna positif ataupun negatif, sehingga *tweet* tersebut tidak akan digunakan lebih lanjut pada tahapan selanjutnya. Dari keseluruhan data yang akan diolah lebih lanjut, diketahui bahwa sebanyak 326 *tweets* memiliki label sentimen positif (label 1) dan 152 *tweets* merupakan sentimen negatif (label 2) seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Demografi Label Data Set

B. Hasil Uji Performa

Setelah dilakukan tahapan *pre-processing*, selanjutnya data set yang sudah bersih dan sudah dilakukan perhitungan frekuensi kemunculan kata menggunakan TF-IDF akan dilakukan pengujian performa menggunakan beberapa algoritma *machine learning*. Pada uji performa ini, adapun yang akan dianalisis dan dibandingkan adalah nilai dari *accuracy*, *precision*, *recall*, dan *f1-score* yang didapatkan dari *confusion matrix*. Pembagian data set yang digunakan adalah 20% data pengujian dan 80% data pelatihan.

Tabel 3. Hasil Uji Performa Naïve Bayes

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0.81	0.91	0.85
Positif	0.95	0.89	0.92
Akurasi	0.90		

Pertama, dilakukan pengujian performa *machine learning* dalam melakukan klasifikasi sentimen menggunakan algoritma Naïve Bayes. Pada Tabel 3 disajikan hasil performa algoritma Naïve Bayes dalam nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Dapat dilihat pada Tabel 3 didapatkan bahwa nilai *precision* untuk label sentimen negatif adalah sebesar 81% dan untuk label sentimen positif adalah sebesar 95%. Nilai *recall* untuk label sentimen negatif adalah sebesar 91% dan untuk label sentimen positif adalah sebesar 89%. Nilai *f1-score* untuk label sentimen negatif adalah sebesar 85% dan untuk label sentimen positif adalah sebesar 92%. Sedangkan untuk nilai akurasi dengan menggunakan Naïve Bayes untuk klasifikasi sentimen masyarakat Indonesia terhadap vaksinasi Covid-19 adalah sebesar 90%.

Tabel 4. Hasil Uji Performa k-Nearest Neighbors (kNN)

	Precision	Recall	F1-Score
Negatif	0.76	1.00	0.86
Positif	0.93	0.37	0.53
Akurasi			0.76

Selanjutnya, dilakukan pengujian performa *machine learning* dalam melakukan klasifikasi sentimen menggunakan algoritma k-Nearest Neighbors (kNN). Pada Tabel 4 disajikan hasil performa algoritma k-Nearest Neighbors (kNN) dalam nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Dapat dilihat pada Tabel 4 didapatkan bahwa nilai *precision* untuk label sentimen negatif adalah sebesar 76% dan untuk label sentimen positif adalah sebesar 93%. Nilai *recall* untuk label sentimen negatif adalah sebesar 100% dan untuk label sentimen positif adalah sebesar 37%. Nilai *f1-score* untuk label sentimen negatif adalah sebesar 86% dan untuk label sentimen positif adalah sebesar 53%. Sedangkan untuk nilai akurasi dengan menggunakan k-Nearest Neighbors (kNN) untuk klasifikasi sentimen masyarakat Indonesia terhadap vaksinasi Covid-19 adalah sebesar 76%.

Tabel 5. Ringkasan Hasil Uji Performa Akurasi

Algoritma	Akurasi
Naïve Bayes	90%
k-Nearest Neighbors (kNN)	76%

Berdasarkan hasil yang ditunjukkan pada Tabel 5 didapatkan bahwa hasil performa akurasi terbaik ditunjukkan oleh algoritma Naïve Bayes dengan nilai akurasi sebesar 90%.

V. KESIMPULAN

Berdasarkan hasil yang didapatkan, dapat disimpulkan bahwa metode pembelajaran mesin dapat digunakan untuk mengklasifikasi sentimen masyarakat Indonesia terhadap program vaksinasi Covid-19 pada sosial media Twitter. Dari seluruh percobaan yang telah dilakukan untuk keseluruhan metode *machine learning* yang telah diuji, algoritma Naïve Bayes menunjukkan performa akurasi tertinggi untuk mengklasifikasikan sentimen masyarakat terhadap program vaksinasi Covid-19 di Indonesia sebesar 90%, dan algoritma k-Nearest Neighbors sebesar 76%.

REFERENSI

[1] APJII (2020). Laporan Survei Internet APJII 2019 – 2020

- [2] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes," *Inti Nusa Mandiri*, vol. 15 no. 1, pp.23-28, 2020
- [3] F. Afshoh, E. W. Pamungkas, "Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter," *Doctoral dissertation, Universitas Muhammadiyah Surakarta*, 2017
- [4] R. Kurniawan, A. Apriliani, "Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper," *Journal uin-alaududin*, vol. 5, no. 1, 2020
- [5] K. Garcia, L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA". *Applied soft computing*, 101: 107057, 2021
- [6] V. Raghupathi, J. Ren, dan W. Raghupathi, "Studying public perception about vaccination: A sentiment analysis of tweets," *International journal of environmental research and public health*, 17(10), 3464, 2020.
- [7] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, dan J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Information*, 12(5), 204, 2021
- [8] M. Ritonga, M. A. Al Ihsan, A. Anjar, dan F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm" *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, p. 012045, IOP Publishing, 2021
- [9] J. Gautam, M. Atrey, N. Malsa, A. Balyan, R. N. Shaw, dan A. Ghosh, "Twitter data sentiment analysis using naïve bayes classifier and generation of heat map for analyzing intensity geographically," *Advances in Applications of Data-Driven Computing* (pp. 129-139), Springer, Singapore, 2021
- [10] A. Muzaki dan A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naïve Bayes Classifier," *Jurnal Teknik Informatika (Jutif)*, vol. 2(2), pp. 101-107, 2021
- [11] A. R. Isnain, J. Supriyanto, dan M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15(2), pp. 121-130.
- [12] F. M. J. M. Shamrat, S. Chakraborty, M. M. Imran, J. N. Muna, M. M. Billah, P. Das, dan O. M. Rahman, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indones. J. Electr. Eng. Comput. Sci.*, 23(1), 2021
- [13] Y. D. Kirana, dan S. Al Faraby, "Sentiment Analysis of Beauty Product Reviews Using the K-Nearest Neighbor (KNN) and TF-IDF Methods with Chi-Square Feature Selection," *Journal of Data Science and Its Applications*, vol. 4(1), pp. 31-42, 2021
- [14] S. Masturoh, dan A. B. Pohan, "Sentiment Analysis Against the Dana E-Wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm," *Pilar Nusa Mandiri: Journal of Computing and Information System*, vol. 17(1), pp. 53-58, 2021