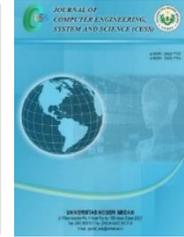


Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Analisis Sentimen Terhadap Vaksin COVID-19 Menggunakan Naive Bayes Classifier, Support Vector Machine dan K-Nearest Neighbors

Sentiment Analysis on COVID-19 Vaccine using Naive Bayes Classifier, Support Vector Machine and K-Nearest Neighbors

Monika Rani*¹, Dian Prawira², Nurul Mutiah³

^{1,2,3} Program Studi Sistem Informasi, Universitas Tanjungpura

Jl. Prof.Dr.H.Hadari Nawawi, Jendral Ahmad Yani, Pontianak - Kalimantan Barat 78124

email: ¹monikarani@student.untan.ac.id, ²dianprawira@sisfo.untan.ac.id, ³nurul@sisfo.untan.ac.id

ABSTRAK

Pengadaan vaksin COVID-19 menyebabkan timbulnya opini yang beragam di masyarakat Indonesia pada media sosial Twitter, opini yang berada di Twitter dapat dilakukan analisis sentimen untuk mengetahui opini masyarakat khususnya pengguna Twitter. Data yang digunakan berupa tweet dengan topik vaksin COVID-19 menggunakan *keyword* vaksin covid 19, vaksin covid, *Astrazeneca*, *Sinovac*, *Moderna*, *Pfizer*, *Novavax* dan *Sinopharm*. Melakukan analisis performa algoritma *Naive Bayes Classifier*, *Support Vector Machine* dan *K-Nearest Neighbors* untuk menentukan hasil tingkat akurasi antara algoritma. Pengujian klasifikasi tertinggi yaitu menggunakan *Support Vector Machine* dengan tingkat akurasi sebesar 0,701. Hasil perbandingan algoritma yang diujikan menggunakan data tweet dengan topik vaksin COVID-19 didapatkan bahwa *Support Vector Machine* lebih baik dari pada antara *Naive Bayes Classifier* dan *K-Nearest Neighbors*. Dari pengujian klasifikasi yang dilakukan menggunakan data tweet vaksin COVID-19 dengan 2500 data. Jumlah data setelah melewati proses pengolahan data menjadi sebanyak 1052 data. Hasil sentimen netral sebanyak 645 sentimen positif sebanyak 250 dan sentimen negatif sebanyak 157.

Kata Kunci: *Naive Bayes Classifier, Support Vector Machine, K-Nearest Neighbors, Analisis Sentimen Twitter, Vaksin COVID-19.*

ABSTRACT

Procurement of the COVID-19 vaccination has led to diverse opinions among Indonesian people on Twitter. Sentiment analysis on Twitter can be carried out to find out public opinion, especially among Twitter users. The data was used in the form of tweets with the topic of the COVID-19 vaccine using the keywords covid 19 vaccine, covid vaccine,

*Penulis Korespondensi:

email: monikarani@student.untan.ac.id

AstraZeneca, Sinovac, Moderna, Pfizer, Novavax and Sinopharm. *analysis of the performance of the Naive Bayes Classifier, Support Vector Machine and K-Nearest Neighbors algorithms to determine the results of the accuracy level between the algorithms.* The highest classification test is using the Support Vector Machine with an accuracy rate of 0.701. The results of the comparison of algorithms tested using tweet data on the topic of the COVID-19 vaccine found that the Support Vector Machine was better than the Naive Bayes Classifier and K-Nearest Neighbors. From the classification test carried out using COVID-19 vaccine tweet data with 2500 data. The amount of data after going through the data processing process is 1052 data. Neutral sentiment results in as many as 645 positive sentiments as many as 250 and negative sentiments as many as 157.

Keywords: *Naive Bayes Classifier, Support Vector Machine, K-Nearest Neighbors, Sentiment Analysis Twitter, COVID-19 Vaccine.*

1. PENDAHULUAN

Wabah penyakit baru yang disebabkan oleh virus corona atau biasa disebut dengan COVID-19 resmi ditetapkan pada tanggal 11 maret 2020 sebagai pandemic global oleh World Health Organization (WHO). Kasus terjangkit virus Corona setiap harinya terus bertambah, pada tanggal 23 mei 2021 lalu, tercatat sebanyak 1.775.220 terkonfirmasi kasus positif COVID-19 dengan jumlah 1.633.045 sembuh dari virus COVID-19 dan 49.328 meninggal. Melihat penyebaran COVID-19 yang terus bertambah serta bahaya yang akan muncul salah satu cara yang dilakukan oleh Kementerian Kesehatan RI untuk mencegah penyebaran virus ini adalah dengan memberikan vaksin COVID-19 dengan total sasaran sebanyak 40.349.049 yang akan diadakan secara bertahap dan diprioritaskan kepada SDM Kesehatan, Petugas publik dan lansia [1].

Vaksin COVID-19 menyebabkan timbulnya opini yang beragam di masyarakat Indonesia Pada media sosial Twitter, Analisis sentimen dilakukan untuk menganalisis opini menggunakan data teks, data yang digunakan berupa tweet dengan topik vaksin COVID-19 menggunakan *keyword* vaksin covid 19, vaksin covid, *AstraZeneca, Sinovac, Moderna, Pfizer, Novavax* dan *Sinopharm* yang di posting pengguna Twitter, data yang didapat akan digunakan untuk melakukan pengujian antar algoritma yaitu *Naive Bayes Classifier, Support Vector Machine* dan *K-Nearest Neighbors*.

Terdapat beberapa algoritma *support vector machine, naive bayes*, serta *K-nearest neighbors* yang digunakan untuk klasifikasi text. Pada algoritma support vector machine bekerja dengan mencari hyperplane terbaik untuk membagi kelas, naive bayes bekerja dengan cara mencari nilai probabilitas bersyarat terbesar dari masing-masing kelas, sedangkan *K-nearest neighbors* bekerja dengan mencari sejumlah k pola yang terdekat [2].

Ketiga algoritma dipilih berdasarkan penelitian sebelumnya, pada penelitian menggunakan algoritma *Support Vector Machine, Naive Bayes* dan *K-nearest neighbors* terhadap terhadap Pandemic Covid-19 diperoleh kesimpulan bahwa algoritma *Support Vector Machine* memiliki akurasi yang lebih tinggi [3]. Pada penelitian menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemic Covid-19 diperoleh *Naive Bayes* memiliki akurasi yang lebih tinggi [4]. Pada penelitian menggunakan algoritma *Naive Bayes* dan KNN terhadap isu COVID-19 diperoleh kesimpulan bahwa algoritma *Naive Bayes* memiliki akurasi lebih tinggi [5].

Pada penelitian ini menguji algoritma yang menghasilkan akurasi yang lebih baik ketika diujikan menggunakan dataset tweet dengan topik vaksin COVID-19. Pengujian menggunakan beberapa algoritma yaitu *Naive Bayes Classifier*, *Support Vector Machine* dan *K-Nearest Neighbors*. Evaluasi perbandingan algoritma menggunakan *confusion matrix* sehingga mendapatkan nilai akurasi dari masing-masing algoritma yang diujikan. Data tweet dengan topik vaksin COVID-19 yang telah didapatkan akan dilakukan visualisasi.

Dengan melakukan analisis sentimen terhadap pengadaan vaksin COVID-19 menggunakan *Naive Bayes Classifier*, *Support Vector Machine* dan *K-Nearest Neighbors* diharapkan mendapatkan hasil sentimen dan hasil akurasi pada masing-masing algoritma yang diujikan. Dari hasil pengujian didapat algoritma yang memiliki nilai akurasi yang lebih baik ketika dilakukan pengujian menggunakan topik vaksin COVID-19.

2. TINJAUAN TEORI

2.1 Text Pre-processing

Text mining bergantung pada berbagai teknik pre-processing untuk mengekstrak representasi terstruktur dari sumber data tidak terstruktur, atau kombinasi keduanya. Teknik pre-processing yang berbeda digunakan bersama-sama untuk membangun representasi dokumen dari data teks asli [6]. Tahapan text pre-processing dalam penelitian ini meliputi:

- 1) *Case Folding* mengubah Text pada dokumen berupa huruf "a" sampai dengan "z" menjadi huruf kecil [7].
- 2) *Cleansing* membersihkan atau menghilangkan dokumen yang berupa angka, url (<http://>), username (@), tanda pagar (#), koma (,), titik (.) dan juga tanda baca lainnya [7].
- 3) *Remove duplicated* merupakan proses untuk menghapus salah satu dokumen yang ditemukan duplikat [7].
- 4) *Replace* merupakan proses untuk melakukan replace atau mengganti kata yang akan diganti [7].
- 5) *Tokenizing* bertugas untuk memotong teks menjadi beberapa bagian yang disebut token [7].
- 6) Normalisasi mengubah kalimat yang ditulis tidak semuanya Bahasa baku, seperti kata "yg", "sbg", "smga" dan lainnya [8].
- 7) *Stopword* mengurangi jumlah kata yang tersimpan dalam daftar token yang nantinya akan diproses seperti kata "di", "yang", "dan", "ke" [9].
- 8) *Stemming* bertujuan untuk mengurangi penambahan suatu imbuhan sehingga menurunkan suatu kata ke bentuk kata dasarnya [7].

2.2 Naive Bayesian Classifiers

Naive bayesian classifiers dapat menghasilkan tingkat akurasi dalam waktu yang cepat Ketika diterapkan menggunakan database yang besar juga pengaruh nilai atribut pada kelas tertentu tidak tergantung pada nilai atribut lainnya. *Naive bayesian classifiers* memiliki tingkat kesalahan minimum dibandingkan dengan pengklasifikasian yang lain, namun dalam penerapannya tidak selalu menghasilkan tingkat kesalahan minimum karena dipengaruhi oleh ketidakakuratan dalam asumsi yang dibuat untuk kurangnya data [10].

Naive Bayesian Classifiers dapat memprediksi probabilitas keanggotaan kelas seperti probabilitas tupel tertentu milik kelas tertentu. Pada *Naive Bayesian Classifiers* efek nilai

atribut pada kelas tidak saling bergantung hal ini disebut *class-conditional independence* dibuat untuk menyederhanakan perhitungan [10].

2.3 Support Vector Machine

Support Vector Machine dapat digunakan untuk data *linear* dan *nonlinear*. Meskipun waktu pelatihan *Support Vector Machine* bisa sangat lambat tetapi menghasilkan akurasi sangat tinggi. *Support Vector Machine* dapat digunakan untuk prediksi numerik serta klasifikasi yang telah diterapkan ke sejumlah bidang, termasuk pengenalan angka tulisan tangan, pengenalan objek, dan identifikasi pembicara [10].

Support Vector Machine dapat digabungkan untuk kasus multi kelas, seperti melatih satu pengklasifikasi per kelas dan penggunaan kode koreksi kesalahan. *Support Vector Machine* dapat digunakan untuk penggunaan data yang besar. *Support Vector Machine* bekerja dengan mencari hyperplane pemisah linear optimal, yaitu batas keputusan yang memisahkan tupel dari satu kelas dari kelas yang lain. *Support Vector Machine* menemukan hyperplane menggunakan vektor pendukung tupel training dan margin oleh vektor pendukung [10].

2.4 K-Nearest Neighbors

K-Nearest Neighbors didasarkan pada pembelajaran dengan analogi yaitu membandingkan tupel pengujian yang diberikan dengan tupel training serupa. Tupel pelatihan dijelaskan oleh ruang pola n-dimensi. *K-Nearest Neighbors* mencari tupel yang tidak diketahui kemudian diberi kelas paling umum di antara *nearest neighbor*. Ketika $k = 1$, tupel yang tidak diketahui ditetapkan kelas dari tupel pelatihan yang paling dekat dengan $k = 1$ ke dalam ruang pola untuk menemukan tupel yang mirip atau paling dekat dengan tupel. *K-Nearest Neighbors* menggunakan perbandingan berbasis jarak yang secara intrinsik menetapkan bobot yang sama untuk setiap atribut. Oleh karena itu mereka dapat menghasilkan akurasi yang buruk ketika diberikan atribut yang tidak relevan [10].

2.5 Confusion Matrix

Setelah dilakukan klasifikasi maka akan dilakukan evaluasi yaitu dengan menggunakan *Confusion Matrix* untuk membandingkan dengan hasil masing-masing klasifikasi *confusion Matrix* merupakan alat ukur yang digunakan untuk mendapatkan jumlah ketepatan dalam proses klasifikasi terhadap kelas dengan menggunakan algoritma yang diujikan [11], pengujian performa *Confusion Matrix* pada Tabel 1.

Tabel 1. *Confusion Matrix*

<i>Confusion Matrix</i>		Nilai sebenarnya	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	<i>TP (True positive)</i> <i>Correct result</i>	<i>FP (False positive)</i> <i>Unexpected result</i>
	<i>False</i>	<i>TN (False Negative)</i> <i>Missing result</i>	<i>TN (True Negative)</i> <i>Correct absence of result</i>

Evaluasi model diukur dengan menggunakan persamaan Akurasi pada, persamaan Presisi pada, persamaan Recall pada dan persamaan F1-Score sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

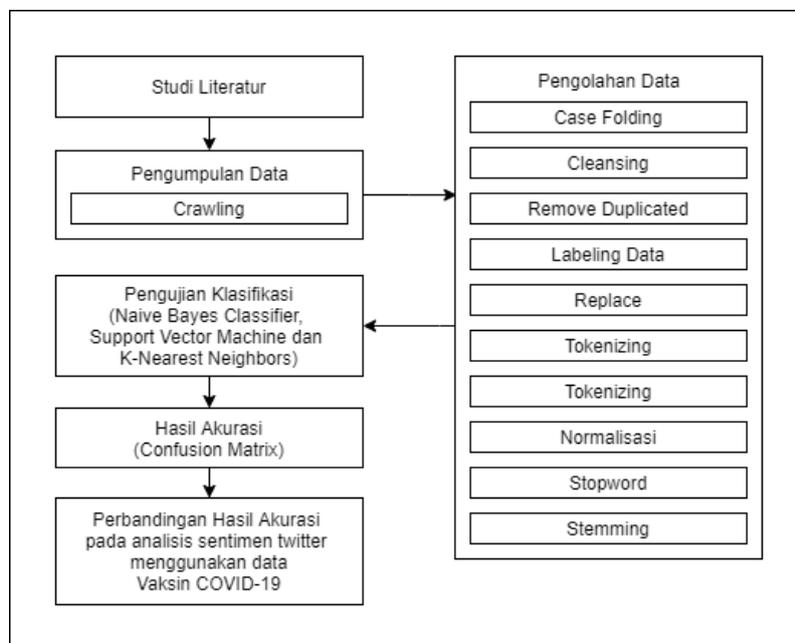
$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

$$\text{f1-Score} = 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

3. METODE

Pada penelitian ini menggunakan *Naive Bayes Classifier*, *Support Machine Vector* dan *K-Nearest Neighbors* untuk melakukan analisis sentimen dan mendapatkan nilai akurasi. Proses yang dilakukan untuk mendapatkan akurasi pada masing-masing algoritma pada Gambar 1.



Gambar 1. Kerangka Penelitian

- 1) Dalam studi literatur peneliti mempelajari bahan-bahan yang relevan dari referensi yang terpercaya untuk mencari informasi yang dibutuhkan.
- 2) Pada tahap pengumpulan data dilakukan proses crawling. Proses crawling dilakukan untuk memperoleh data dari media sosial Twitter untuk melakukan penelitian. Pengambilan data menggunakan library Python Tweepy dan Twitter API dengan topik Vaksin COVID-19.
- 3) Pada tahap pengolahan data terdiri dari *case folding*, *cleansing*, *remove duplicated*, *labelling data*, *replace*, *tokenization*, *normalisasi*, *stopword* dan *stemming*
- 4) Setelah melakukan tahap pengolahan data akan dilakukan klasifikasi dengan mengujiikan menggunakan *Naive Bayes Classifier*, *Support Machine Vector* dan *K-Nearest Neighbors*, dari ketiga algoritma.
- 5) Evaluasi model dengan menggunakan *confusion matrix*.

- 6) Hasil nilai akurasi yang didapat akan disimpulkan algoritma yang menghasilkan nilai akurasi lebih tinggi dari algoritma yang diujikan.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan pada tahap awal untuk pengumpulan data tweet menggunakan topik vaksin COVID-19 berupa *keyword* vaksin covid 19, vaksin covid, *AstraZeneca*, *Sinovac*, *Moderna*, *Pfizer*, *Novavax* dan *Sinopharm* menggunakan Twitter API dengan menggunakan *library tweepy* untuk mengakses Twitter API. Penelitian menggunakan kata kunci vaksin covid 19, vaksin covid, *AstraZeneca*, *Sinovac*, *Moderna*, *Pfizer*, *Novavax* dan *Sinopharm* dengan jumlah data sebanyak 2500 yang diambil pada rentang waktu dari 2022-03-01 sampai 2022-03-08, seperti pada Tabel 2.

Tabel 2. Hasil Pengumpulan Data

User	Tanggal	Tweet
DindaNatasha15	2022-03-07 23:56:48	Pemerintah sudah pastikan bahwa vaksin covid 19 sudah terbukti aman dan teruji. Patuhi Disiplin Prokes https://t.co/c4BOiRwLyW
eknug	2022-03-07 23:22:42	"Sudah Vaksin Booster, Gibran Positif Covid-19 Lagi https://t.co/aXbDnWsJHe "
nengsunshine	2022-03-07 23:47:55	"Ok Google! Kalau sudah negatif dari Covid, kapan vaksin bisa dilakukan?"

4.2 Pengolahan Data

Tahap pengolahan data dilakukan untuk membuat data tweet yang telah diambil menjadi kata yang baku menggunakan Bahasa Indonesia serta mengurangi data tweet dari kata yang tidak diperlukan. Adapun proses yang dilakukan sebagai berikut:

- 1) *Case folding* dilakukan untuk mengubah semua huruf besar berubah menjadi huruf kecil, seperti pada Tabel 3.

Tabel 3. Proses *Case Folding*

Sebelum	Sesudah
Patuhi Disiplin Prokes https://t.co/c4BOiRwLyW	patuhi disiplin prokes https://t.co/c4boirwlyw

- 2) *Cleansing* dilakukan untuk menghilangkan tanda baca serta tautan seperti pada Tabel 4.

Tabel 4. Proses *Cleansing*

Sebelum	Sesudah
patuhi disiplin prokes https://t.co/c4boirwlyw	patuhi disiplin prokes

- 3) *Remove Duplicated* dilakukan untuk menghilangkan data ditemukan duplikat, Setelah melewati proses *Remove Duplicated* terjadi perubahan jumlah data yang awalnya data berjumlah 2500 menjadi 1052 data seperti pada Tabel 5.

Tabel 5. Proses Remove Duplicated

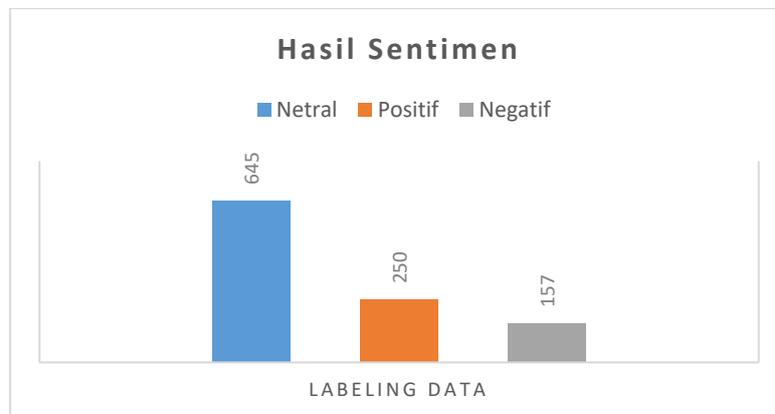
Sebelum	Sesudah
patuhi disiplin prokes	patuhi disiplin prokes
patuhi disiplin prokes	

- 4) *labeling* data dilakukan untuk mengategorikan tweet ke dalam kategori positif, netral dan negatif, seperti pada Tabel 6.

Tabel 6. Proses Labelling Data

Tweet	Sentimen
pemerintah sudah pastikan bahwa vaksin covid 19 sudah terbukti aman dan teruji patuhi disiplin prokes	Positif
sudah vaksin booster Gibran positif covid 19 lagi	Negatif
ok google kalau sudah negatif dari covid kapan vaksin bisa dilakukan	Netral

Pemberian label dilakukan pada data sebanyak 1052, terdapat tweet dengan sentimen netral sebanyak 645, positif sebanyak 250 dan negatif sebanyak 157. Gambar hasil sentimen pada Gambar 2.



Gambar 2. Hasil Sentimen

- 5) *replace* dilakukan untuk mempertahankan kalimat kedua pada "vaksin kedua" agar tidak terhapus pada proses selanjutnya, seperti pada Tabel 7.

Tabel 7. Proses Replace

Sebelum	Sesudah
Jadwal vaksin kedua dilaksanakan	Jadwal vaksinkedua dilaksanakan

- 6) *tokenizing* dilakukan untuk memecahkan kalimat atau teks menjadi beberapa bagian yang disebut dengan token, seperti pada Tabel 8.

Tabel 8. Proses *Tokenizing*

Sebelum	Sesudah
patuhi disiplin proses	['patuhi', 'disiplin', 'proses']

- 7) normalisasi dilakukan untuk mengubah kalimat yang tidak baku atau singkatan menjadi kalimat baku, seperti pada Tabel 9.

Tabel 9. Proses Normalisasi

Sebelum	Sesudah
['disiplin', ' prokes ']	['disiplin', 'protokol kesehatan']

- 8) *stopword* dilakukan untuk menghilangkan kata penghubung, seperti pada Tabel 10.

Tabel 10. Proses *Stopword*

Sebelum	Sesudah
['sudah', 'terbukti', 'aman', ' dan ', 'terbukti', 'aman', 'teruji']	['terbukti', 'aman', 'teruji']

- 9) *stemming* dilakukan untuk menghilangkan kata imbuhan menjadi kata dasar, seperti pada Tabel 11.

Tabel 11. Proses *Stemming*

Sebelum	Sesudah
[' terbukti ', 'aman', ' teruji ']	bukti aman uji

4.3 Pengujian Klasifikasi

Data tweet yang sudah melewati pre-processing dan juga labelling akan dilakukan pembobotan kata. Pada proses pembobotan menggunakan TF-IDF, kata akan dikonversikan menjadi numerik atau angka. Setelah proses pembobotan TF-IDF dilakukan pembagian data terlebih dahulu dengan total data 1052 akan dibagi menjadi data train dan data test, data train sebesar 80% dengan jumlah 841 data dan data test sebesar 20% dengan jumlah 211 data.

4.3.1 Perhitungan *Naive Bayes Classifier*

Perhitungan performa *Naive Bayes Classifier* menggunakan *Confusion Matrix* yang telah dilakukan terdapat data yang benar sebesar 19 data negatif, 85 data netral dan 40 data positif.

Tabel 12. Hasil Perhitungan *Naive Bayes Classifier*

		Akurasi= 0,682			Presisi	Recall	f-1 Score
<i>confusion matrix</i>		Kelas prediksi					
		Negatif	Netral	Positif			
Kelas sebenarnya	Negatif	19	16	2	0,513	0,432	0,468
	Netral	24	85	9	0,720	0,732	0,725
	Positif	1	15	40	0,714	0,784	0,746

Pada Tabel 12 merupakan hasil dari evaluasi model, akurasi merupakan persentase dari jumlah data relevan sebesar 0,682. Presisi merupakan nilai perbandingan antara jumlah data relevan dengan jumlah data, untuk negatif sebesar 0,513, untuk netral sebesar 0,720 dan untuk positif sebesar 0,714. Nilai *recall* merupakan keberhasilan menemukan data relevan, untuk negatif sebesar 0,432, untuk netral sebesar 0,732 dan untuk positif sebesar 0,784. Nilai *f1-Score* merupakan pembandingan untuk mengukur data rata-rata dengan menggabungkan *presisi* dan *recall*, untuk negatif sebesar 0,468, untuk netral sebesar 0,725 dan untuk positif sebesar 0,746.

4.3.2 Perhitungan Support Machine Vector

Perhitungan performa *Support Vector Machine* menggunakan *Confusion Matrix* yang telah dilakukan terdapat data yang benar sebesar 18 data negatif, 94 data netral dan 36 data positif.

Tabel 13. Hasil Perhitungan Support Vector Machine

<i>confusion matrix</i>		Akurasi=0,701			Presisi	Recall	f-1 Score
		Kelas prediksi					
		Negatif	Netral	Positif			
Kelas	Negatif	18	12	3	0,545	0,409	0,465
sebenarnya	Netral	26	94	12	0,712	0,814	0,758
	Positif	0	10	36	0,782	0,705	0,741

Pada Tabel 13 merupakan hasil dari evaluasi model, akurasi merupakan persentase dari jumlah data relevan sebesar 0,701. Presisi merupakan nilai perbandingan antara jumlah data relevan dengan jumlah data, untuk negatif sebesar 0,545, untuk netral sebesar 0,712 dan untuk positif sebesar 0,782. Nilai *recall* merupakan keberhasilan menemukan data relevan, untuk negatif sebesar 0,409, untuk netral sebesar 0,814 dan untuk positif sebesar 0,705. Nilai *f1-Score* merupakan pembandingan untuk mengukur data rata-rata dengan menggabungkan *presisi* dan *recall*, untuk negatif sebesar 0,465, untuk netral sebesar 0,758 dan untuk positif sebesar 0,741.

4.3.3 Perhitungan K-Nearest Neighbors

Perhitungan performa *K-Nearest Neighbors* menggunakan *Confusion Matrix* yang telah dilakukan terdapat data yang benar sebesar 28 data negatif, 85 data netral dan 30 data positif.

Tabel 14. Hasil Perhitungan K-Nearest Neighbors

<i>Confusion matrix</i>		Akurasi=0,677			Presisi	Recall	f-1 Score
		Kelas prediksi					
		Negatif	Netral	Positif			
Kelas	Negatif	28	28	11	0,417	0,636	0,503
sebenarnya	Netral	16	85	10	0,765	0,732	0,746
	Positif	0	3	30	0,909	0,731	0,713

Pada Tabel 14 merupakan hasil dari evaluasi model, tingkat akurasi merupakan persentase dari jumlah data relevan sebesar 0,677. Presisi merupakan nilai perbandingan antara jumlah data relevan dengan jumlah data, untuk negatif sebesar 0,417, untuk netral sebesar 0,765 dan untuk positif sebesar 0,909. Nilai *recall* merupakan keberhasilan menemukan data relevan, untuk negatif sebesar 0,636, untuk netral sebesar 0,732 dan untuk positif sebesar 0,731. Nilai *f1-Score* merupakan perbandingan untuk mengukur data rata-rata dengan menggabungkan *presisi* dan *recall*, untuk negatif sebesar 0,503, untuk netral sebesar 0,746 dan untuk positif sebesar 0,713.

Tabel 15. Hasil Pengujian

Metode klasifikasi				
Akurasi=0,682				
<i>Naive Bayes Classifier</i>	Kelas prediksi	Presisi	Recall	F1-Score
	Negatif	0,513	0,432	0,468
	Netral	0,720	0,732	0,725
	Positif	0,714	0,784	0,746
Akurasi=0,701				
<i>Support Vector Machine</i>	Kelas prediksi	Presisi	Recall	F1-Score
	Negatif	0,545	0,409	0,465
	Netral	0,712	0,814	0,758
	Positif	0,782	0,705	0,741
Akurasi=0,677				
<i>K-Nearest Neighbors</i>	Kelas prediksi	Presisi	Recall	F1-Score
	Negatif	0,417	0,636	0,503
	Netral	0,765	0,732	0,746
	Positif	0,909	0,731	0,713

Tabel 15 merupakan hasil perbandingan tingkat akurasi antara *Naive Bayes Classifier*, *Support Vector Machine* dan *K-Nearest Neighbors* pada pengujian yang dilakukan dengan menggunakan 1052 data kemudian dibagi menjadi data train 80% dengan jumlah 833 data dan data test 20% dengan jumlah 211 data. Menghasilkan tingkat akurasi pada masing-masing algoritma yang diujikan pada *Naive Bayes Classifier* tingkat akurasi sebesar 0,682, pada *Support Vector Machine* tingkat akurasi sebesar 0,701 dan *K-Nearest Neighbors* tingkat akurasi sebesar 0,677. Maka dapat disimpulkan bahwa *Support Vector Machine* merupakan algoritma yang lebih cocok untuk klasifikasi data tweet vaksin COVID-19.

5. KESIMPULAN

Pada penelitian ini mendapatkan hasil perbandingan tingkat akurasi tertinggi yaitu *Support Vector Machine* mendapatkan tingkat akurasi sebesar 0,701, *Naive Bayes Classifier* mendapatkan tingkat akurasi sebesar 0,682 dan *K-Nearest Neighbors* mendapatkan tingkat akurasi sebesar 0,677. *Support Vector Machine* mampu mencapai tingkat akurasi yang tinggi karena bekerja dengan meningkatkan dimensi data pelatihan. Pada penelitian ini mendapatkan respon terhadap pemberian vaksin COVID-19 pada tweet dengan hasil

sentimen netral sebanyak 645 sentimen, positif sebanyak 250 dan sentimen negatif sebanyak 157.

REFERENSI

- [1] Kementerian Kesehatan RI, "Update Perkembangan Vaksinasi COVID-19 Di Indonesia, Per Tanggal 23 Mei 2021 Pukul 18.00 WIB," 2021. [Online]. Available: https://www.facebook.com/story.php?story_fbid=4527156370647761&id. [Diakses 23 Mei 2021].
- [2] K. Fithriasari, R. W. Mayasari, N. Iriawan dan W. S. Winahju, "Surabaya Government Performance Evaluation Using Tweet Analysis," *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, pp. 31-42, 2022.
- [3] F. S. Pamungkas dan I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *In PRISMA, Prosiding Seminar Nasional Matematika*, pp. 628-634), 2021.
- [4] B. Laurensz dan E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *urnal Nasional Teknik Elektro dan Teknologi Informasi*, pp. 118-123, 2021.
- [5] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn," *Inti Nusa Mandiri*, pp. 23-28, 2020.
- [6] R. Feldman dan J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*, The text mining handbook: advanced approaches in analyzing unstructured data, 2007.
- [7] M. N. Rizaldi, A. Adiwijaya dan S. A. Faraby, "Klasifikasi Argument Pada Teks dengan Menggunakan Metode Multinomial Logistic Regression Terhadap Kasus Pemindahan Ibu Kota Indonesia di Twitter," *Jurnal Media Informatika Budidarma 4*, vol. 4(4), pp. 904-913, 2020.
- [8] S. Khairunnisa, . A. Adiwijaya dan S. A. Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19," *Jurnal Media Informatika Budidarma*, pp. 406-414, 2021.
- [9] M. I. Aditama, R. I. Pratama, K. H. U. Wiwaha dan N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," *Journal Information Engineering and Educational Technology) ISSN 2549*, 2022.
- [10] J. Han, J. Pei dan M. Kamber, *Data Mining: Concepts And Techniques*, Elsevier, 2012.
- [11] O. Arifin dan T. B. Sasongko, "Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine Dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA," *SEMNASTEKNOMEDIA ONLINE*, vol. 6(1), pp. 1-2, 2018.