

Analisis Sentimen untuk Program Vaksin *Booster* Sebagai Syarat Mudik 2022 Menggunakan Algoritma KNN

Sentiment Analysis for The Booster Vaccine Program as A Condition for Homecoming 2022 Using The KNN Algorithm

Andreansyah^{1*}, Grace Gata²

^{1,2}Fakultas Teknologi Informasi
Universitas Budi Luhur
Email: ¹andreoco27@gmail.com, ²grace.gata@budiluhur.ac.id
(* : corresponding author)

Abstract

Social media as a place to access and disseminate information has increased, one of which is Twitter. Twitter in this study is used as a source to find public opinion on government policies regarding booster vaccines as a condition for going home in Indonesia in 2022. These opinions can be positive or negative, depending on the public's view of the object. So the purpose of this study is to conduct a sentiment analysis using the K-Nearest-Neighbors algorithm to determine how much accuracy is generated from the K-Nearest Neighbors algorithm in classifying the sentiment of a tweet related to booster vaccines as a condition for going home. This study uses the Cross-Industry Standard Process For Data Mining (CRISP-DM) methodology. By going through several stages, such as preprocessing and modeling steps. The data obtained is 2,384 with positive sentiment 1,970 tweets and negative views 414 tweets. Based on the results obtained from the modeling stage using the K-Nearest Neighbors algorithm for split data modeling comparison 70:30 using k=9, the accuracy obtained is 85.17%, precision is 85.51%, and recall is 98.82%.

Keywords : *sentiment analysis, text mining, k-nearest neighbors, homecoming, vaccine booster.*

Abstrak

Media sosial sebagai tempat untuk mengakses dan menyebarkan informasi telah berkembang sangat pesat, salah satunya adalah *twitter*. *Twitter* pada penelitian ini digunakan sebagai sumber untuk mengetahui pendapat atau opini masyarakat terhadap kebijakan pemerintah tentang vaksin *booster* sebagai syarat mudik di Indonesia tahun 2022. Opini-opini tersebut bisa berupa opini positif atau negatif, tergantung dari pandangan publik terhadap objek tersebut. Maka tujuan dari penelitian ini adalah untuk melakukan analisis sentimen dengan menggunakan algoritma *K-Nearest-Neighbors*, untuk mengetahui seberapa besar akurasi yang dihasilkan dari algoritma *K-Nearest Neighbors* dalam mengklasifikasikan sentimen dari sebuah *tweet* yang berkaitan dengan vaksin *booster* sebagai syarat mudik. Pada penelitian ini menggunakan metodologi *Cross-Industry Standard Process For Data Mining* (CRISP-DM). Dengan melewati beberapa tahapan seperti tahapan *preprocessing* dan pemodelan. Jumlah data yang didapat 2384 dengan sentimen positif 1970 *tweet* dan sentimen negatif 414 *tweet*. Berdasarkan hasil yang didapat dari tahapan pemodelan dengan menggunakan algoritma *K-Nearest Neighbors* pemodelan *split data* perbandingan 70:30 menggunakan k=9, maka *accuracy* yang dihasilkan sebesar 85.17%, *precision* 85.51% dan *recall* 98.82%.

Kata Kunci : *analisis sentimen, text mining, k-nearest neighbors, mudik, vaksin booster.*

1. PENDAHULUAN

Wabah *Covid-19* ditetapkan secara resmi sebagai pandemi global oleh *World Health Organization* (WHO) pada tanggal 11 Maret 2020. Penyebaran virus tersebut dimulai pada akhir tahun 2019. Masa Pandemi *Covid-19* di tahun ke-3 saat ini masih terus dilakukan waspada bagi seluruh masyarakat [1].

Melihat pesatnya penyebaran *Covid-19* dan dampak yang akan ditimbulkan apabila tidak secepatnya diselesaikan, salah satu solusi untuk mengurangi tingkat penyebaran virus yaitu dengan Vaksin *Covid-19*. Vaksin *Covid-19* melalui program vaksin dosis 1 dan 2, serta saat ini terus dilakukan pelayanan pemberian Vaksin *Booster Covid-19* ketiga yang bias didapatkan 6 (enam) bulan setelah dilakukan vaksin kedua. Vaksin akan melindungi masyarakat terhadap penyebaran virus dan juga dapat menurunkan tingkat penyebaran virus tersebut [2].

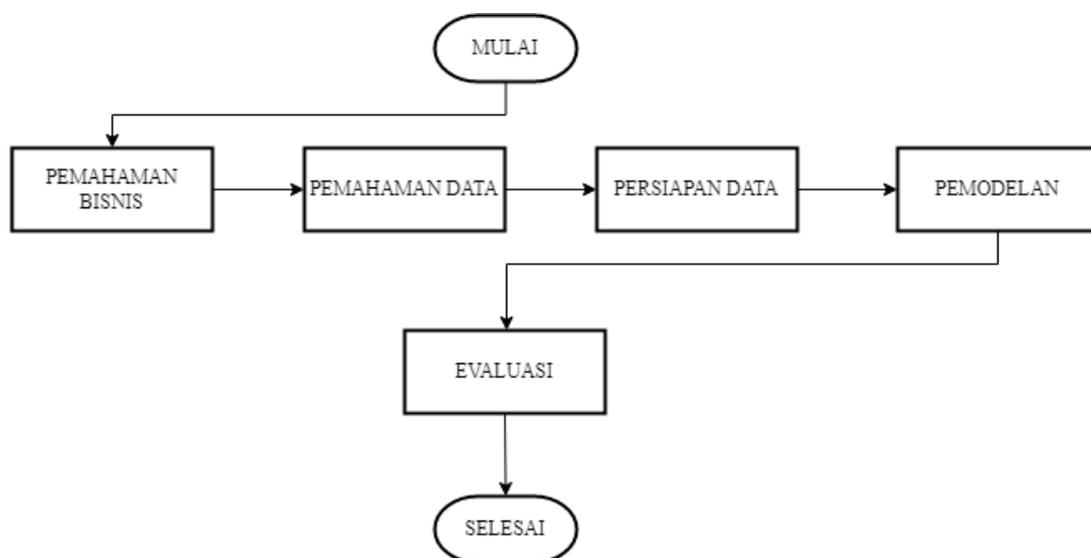
Vaksin *Booster Covid-19* ketiga juga menjadi salah satu syarat mudik di tahun 2022. Mudik adalah kebiasaan masyarakat yang biasa dilakukan pada saat menjelang Idul Fitri. Mudik juga merupakan kegiatan perjalanan dari kota ke desa yang dilakukan secara serentak atau masal, selain dari pertimbangan budaya dan agama[3].

Namun banyak masyarakat yang setuju dan tidak setuju terhadap Vaksin *Booster Covid-19* ketiga sebagai syarat mudik 2022. Sehingga hal tersebut menjadi pembicaraan diberbagai media sosial, salah satunya media sosial *Twitter*. Opini masyarakat di media sosial *Twitter* pun sangat beragam, sehingga terdapat respon yang berbeda dari setiap masyarakat. Kemungkinan ada respon dari masyarakat yaitu respon positif maupun negatif [4]. Respon masyarakat dapat dilihat pada beberapa *tweet* dengan menggunakan *keyword* “Vaksin *Booster*” dan “Mudik”. Sehingga, dibutuhkan penelitian sentimen analisis terhadap opini-opini tersebut dengan tujuan menjadikan referensi bagi pemerintah atau pihak berkepentingan lainnya dalam membuat peraturan mudik di tahun 2022 atau tahun berikutnya [5]

Penelitian ini menggunakan metode *K-Nearest Neighbors* (KNN). KNN adalah salah satu algoritma yang mudah diimplementasikan dengan tingkat keefektifan yang tinggi, serta cocok untuk berbagai masalah yang berhubungan dengan klasifikasi [6]. Maka tujuan dari penelitian ini adalah melakukan analisis sentimen dengan menggunakan algoritma KNN dalam mengklasifikasikan sentimen dari sebuah *tweet* yang berkaitan dengan vaksin *booster* sebagai syarat mudik 2022.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1: Tahapan Penelitian Analisis Sentimen untuk Program Vaksin *Booster* Sebagai Syarat Mudik 2022 menggunakan Algoritma KNN

Pada penelitian ini terdiri dari beberapa tahapan yang ditunjukkan pada gambar 1. Penelitian ini menerapkan metodologi *Cross-Industry Standard Process For Data Mining* (CRISP-DM). Berikut adalah penjelasan mengenai tahapan-tahapan yang dilakukan selama penelitian [7]:

a. Pemahaman Bisnis

Pada tahapan ini peneliti memahami permasalahan yang diangkatnya yaitu analisis sentimen terhadap vaksin *booster* sebagai syarat mudik 2022, penelitian ini dilakukan dengan mempelajari artikel atau berita di internet tentang vaksin *booster* sebagai syarat mudik 2022, kemudian menggali informasi

dan mencari data *tweet* yang bersumber dari media sosial *twitter* yang berkaitan dengan vaksin *booster* dan mudik, dari postingan *tweet* tersebut bisa mengetahui komentar dari masyarakat.

b. Pemahaman Data

Pada tahapan ini peneliti memproses pengambilan data melalui *twitter* dengan menggunakan *keywords* seperti vaksin *booster* dan mudik. Proses pengambilan data pada penelitian ini menggunakan *software rapid miner*. Sebelum melakukan proses pengambilan *tweet*, peneliti diwajibkan untuk membuat koneksi yang menghubungkan antara *rapid miner* dengan *twitter* agar proses pengumpulan data *tweet* dapat dilakukan. Data yang dikumpulkan dari tanggal 27 Mei 2022 hingga 30 Mei 2022. Dengan data *tweet* yang dikumpulkan sebanyak 8341 *tweet*. Data yang diolah adalah data yang sesuai dengan *keywords* sebelumnya yang telah disesuaikan. Setelah mengumpulkan data yang berkaitan dengan vaksin *booster* dan mudik, kemudian data digabung menjadi satu, lalu peneliti melakukan *remove duplicate* menggunakan *rapid miner* untuk menghapus data yang muncul secara berulang. Setelah melakukan *remove duplicate* maka sebelum melakukan *preprocessing* data, data perlu diberi label sentimen terlebih dahulu, pelabelan data dilakukan dengan menggunakan metode *lexicon*.

c. Persiapan Data

Pada tahapan ini data yang sudah didapatkan sebelumnya, akan ditentukan sentimennya. Setelah pelabelan sentimen dilakukan, maka data *tweet* yang sudah memiliki label sentimen akan dilakukan data *preprocessing*. Tahapan-tahapan saat melakukan *preprocessing*[8]:

- 1) *Remove Duplicate* adalah proses untuk menghapus data yang muncul secara berulang.
- 2) *Replace* adalah proses untuk menghilangkan *stopwords* yang terdapat pada *tweet* seperti RT, https:, dan @username.
- 3) *Case Folding* adalah operator yang berfungsi untuk mengubah semua huruf menjadi huruf kecil.
- 4) *Tokenizing* adalah proses untuk memecah data *tweet* yang masih dalam bentuk kalimat menjadi individual.
- 5) *Filter Tokens (By Length)* adalah operator untuk memfilter kata berdasarkan panjang karakter.
- 6) *Stem (Dictionary)* adalah proses mengembalikan kata menjadi bentuk kata dasar dari sebuah kata.
- 7) *Stopwords (Dictionary)* adalah proses yang dilakukan untuk menghilangkan kata sambungan.
- 8) *Term Weighting (TF-IDF)* adalah proses untuk melakukan pembobotan kata, dengan cara mencari nilai *Term Frequency (TF)*, lalu mencari nilai *Document Frequency (DF)*, kemudian mencari nilai *Invers Document Frequency (IDF)*, dan menghitung bobot[9].

d. Pemodelan

Pada tahapan ini peneliti akan melakukan pemodelan dengan menggunakan dua tahapan yaitu *split data* dan *cross fold validation*. Pada *split data* membagi dataset tersebut dengan perbandingan data *training* dan data *testing* yaitu 60:40, 70:30, dan 80:20. Lalu pada tahap *Cross Validation* dengan pembagian secara acak kedalam 10 bagian (*number of folds* =10). Pada tahap pemodelan ini menggunakan algoritma KNN namun sebelumnya telah dilakukan perbandingan model klasifikasi dengan menggunakan algoritma *Naïve Bayes* dan SVM[10].

e. Evaluasi

Pada tahapan ini peneliti akan melakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan *confusion matrix* terhadap algoritma *K-Nearest Neighbors*.

3. HASIL DAN PEMBAHASAN

3.1 Pemahaman Terhadap Data

Pada tahapan ini peneliti melakukan pengumpulan data dari media sosial *twitter* dengan menggunakan *rapidminer* serta peneliti akan memahami data yang akan digunakan, kemudian peneliti menentukan *keywords* yang berkaitan dengan vaksin *booster* sebagai syarat mudik 2022 untuk proses *crawling data tweet*. Peneliti juga melakukan pengambilan data pada tanggal 27 Mei 2022 hingga 30 Mei 2022. Berikut ini adalah data *tweet* yang berhasil peneliti kumpulkan dari *twitter* dengan memakai *tools rapidminer* seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Jumlah Data *Tweet* Hasil *Crawling*

Keywords	Tanggal	Jumlah
Vaksin Booster	27 – 30 Mei 2022	4536
Mudik	27 – 30 Mei 2022	3895
Jumlah		8.431

3.2 Persiapan Data

Pada tahapan persiapan data ini peneliti akan membagi menjadi 2 tahapan. Pada tahapan *preprocessing* proses yang akan dilakukan oleh peneliti setelah data *tweet* sudah memiliki sentimen positif atau negatif dan tahapan penentuan data latih dan data uji, tahapan ini merupakan proses menentukan perbandingan data latih dan data uji terlebih dahulu sebelum masuk ke dalam tahapan pemodelan.

3.2.1 Tahapan *Preprocessing*

Membersihkan data dilakukan dengan menggunakan tahapan *preprocessing* yang terdiri dari *tokenize*, *transform cases*, *filter tokens (by Length)*, *stem (Dictionary)* dan *filter stopwords (Dictionary)*. Dimana dalam penggunaan *filter stopwords* peneliti menggunakan kamus *stopwords* milik *Kaggle.com* dan penggunaan *stem* peneliti membuat manual dengan bantuan kamus besar bahasa Indonesia. Simpan kata-kata yang ada dalam dataset dalam bentuk *file notepad* atau *.txt* yang memproses dataset dengan *rapid miner* yang selanjutnya membuat kalimat menjadi kata dasar. Berikut ini adalah contoh hasil dari tahapan *preprocessing* Tabel 2.

Tabel 2. Perbandingan Sebelum dan Sesudah *Preprocessing*

Preprocessing	Sebelum	Sesudah
<i>Transform Cases</i>	Presiden Jokowi Minta Masyarakat Tak Pilihpilih Jenis Vaksin <i>Booster</i>	presiden jokowi minta masyarakat tak pilihpilih jenis vaksin <i>booster</i>
<i>Tokenize</i>	Jokowi Pastikan Stok Vaksin <i>Booster Covid RI</i> Lebih dari Cukup	Jokowi Pastikan Stok Vaksin Booster Covid RI Lebih dari Cukup
<i>Filter Tokens (by Length)</i>	Alhamdulillah di mudik kemarin berlangsung aman dan sehat tanpa ada peristiwa yang berakibat fatal serta tidak ada penambahan kasus <i>Covid19</i> yang berarti ujar Presiden	Alhamdulillah mudik kemarin berlangsung aman dan sehat tanpa ada peristiwa yang berakibat fatal serta tidak penambahan kasus <i>Covid19</i> berarti ujar presiden
<i>Filter Stopwords (Dictionary)</i>	Alhamdulillah mudik kemarin berlangsung aman dan sehat tanpa ada peristiwa yang berakibat fatal serta tidak penambahan kasus <i>Covid19</i> berarti presiden	Alhamdulillah mudik kemarin aman dan sehat tanpa ada peristiwa yang berakibat fatal serta penambahan <i>Covid19</i> presiden
<i>Stem (Dictionary)</i>	Alhamdulillah mudik kemarin sehat tanpa peristiwa berakibat fatal serta penambahan <i>Covid19</i> presiden	Alhamdulillah mudik kemarin sehat tanpa peristiwa akibat fatal serta tambah <i>Covid19</i> presiden

3.2.2 Pembagian Data Training dan Data Testing

Data latih adalah data yang akan digunakan untuk proses latihan bagi model klasifikasi. Sedangkan data uji adalah data yang akan digunakan untuk uji *rule* klasifikasi. Pada penelitian ini data akan dibagi menjadi perbandinagn 60:40, 70:30 dan 80:20 yang dapat dilihat pada Tabel 3. Metode pembagiannya menggunakan metode *stratified sampling*.

Tabel 3. Sebaran Data *Training* dan Data *Testing* Pada dataset (ulasan masyarakat terhadap vaksin *booster* syarat mudik)

Perbandingan 60:40			
Sentimen	Training	Testing	Total
Positive	1182	788	1970
Negative	248	166	414
Total	1430	954	2384
Perbandingan 70:30			
Sentimen	Training	Testing	Total
Positive	1379	591	1970
Negative	290	124	414
Total	1669	715	2384
Perbandingan 80:20			
Sentimen	Training	Testing	Total
Positive	1576	394	1970
Negative	331	83	414
Total	1907	477	2384

3.3 Pemodelan

Pada tahapan Pemodelan ini, penelitian ini akan dilakukan pengukuran performa klasifikasi dengan dua tahap:

- Pemodelan memakai *split data*. Pada tahapan ini data akan di *split* menjadi 3 rasio yaitu 60:40, 70:30, dan 80:20. Detail dataset dapat dilihat pada Tabel 3.
- Pemodelan menggunakan *cross validation*. Pada tahapan ini data *training* dan data *testing* di *split* secara acak menjadi 10 bagian (*number of folds* = 10).

3.3.1 Hasil Pemodelan Menggunakan *Split Data*

Pada tabel 4 adalah hasil pemodelan klasifikasi menggunakan *Split Data* dengan melakukan perbandingan antara algoritma KNN, *Naïve Bayes* dan SVM. Berdasarkan tampilan pada tabel 4, dapat disimpulkan bahwa nilai akurasi tertinggi ada pada algoritma KNN dengan akurasi sebesar 85.17% dibandingkan dengan algoritma *Naïve Bayes* dan SVM.

Tabel 4. Perbandingan Akurasi Algoritma KNN, *Naïve Bayes* dan SVM Menggunakan Metode *Split Data*

Perbandingan	Algorithm		
	KNN	Naïve Bayes	SVM
60:40	84.80%	63.73%	82.70%
70:30	85.17%	63.78%	82.52%
80:20	84.91%	62.89%	82.39%

3.3.2 Hasil Proses Pemodelan Menggunakan *Cross Validation*

Pada tabel 5 adalah hasil pemodelan klasifikasi menggunakan *Cross Validation* dengan melakukan perbandingan antara algoritma KNN, *Naïve Bayes* dan SVM. Berdasarkan tampilan pada tabel 5, dapat disimpulkan bahwa nilai akurasi tertinggi ada pada algoritma KNN dengan akurasi sebesar 84.48% dibandingkan dengan algoritma *Naïve Bayes* dan SVM.

Tabel 5. Hasil akurasi Algoritma KNN, *Naïve Bayes* dan SVM Menggunakan Metode *Cross Validation*

Algoritma	Accuracy
<i>K-Nearest Neighbors</i>	84.48%
<i>Naïve Bayes</i>	62.75%
<i>Support Vector Machine</i>	83.18%

Berdasarkan Tabel 4. dan Tabel 5. yang berisi hasil pengukuran performa model klasifikasi *K-Nearest Neighbors* dengan 2 cara yaitu pembagian dataset dengan *split data* dan *cross validation*, maka dapat disimpulkan bahwa pemodelan dengan *split data* 70:30 memiliki akurasi sebesar 85.17% sedangkan dengan pemodelan *cross validation* memiliki akurasi sebesar 84.48%, sehingga akurasi tertinggi yang dihasilkan pada penelitian ini adalah dengan pemodelan dengan *split data* 70:30.

3.3.3 Hasil Perhitungan Algoritma K-Nearest Neighbors

Perhitungann manual dengan algoritma *K-Nearest Neighbors* ini menggunakan *sample* dari dataset yang diambil secara acak sebanyak 5 data *tweet* yang dapat dilihat pada Tabel 6. yang dimana *sample tweet* D1 sampai D4 sebagai data *training* dan D5 sebagai data *testing*.

Tabel 6. Contoh *Tweet*

<i>Sample</i>	<i>Tweet</i>	<i>Sentimen</i>
D1	Bismillah besok vaksin <i>booster</i>	<i>Positive</i>
D2	Booster Jadi Syarat Mudik Upaya Pemerintah Untuk Melindungi Masyarakat	<i>Positive</i>
D3	Bisa lemes sama pusing gini efek vaksin <i>booster</i>	<i>Negative</i>
D4	GES TANGANKU SAKIT BGT HABIS VAKSIN <i>BOOSTER</i> HIKS GABISA NAIK TURUN LENGANNYA SAKIT	<i>Negative</i>
D5	Apa bener syarat mudik harus udah vaksin <i>booster</i>	?

Setelah menentukan *sample tweet* yang ingin digunakan, maka tahapan berikutnya adalah menentukan parameter K, kemudian melakukan *preprocessing*, dan menentukan bobot *term* dari 5 *tweet* menggunakan *Term Weighting* (TF-IDF). Pada contoh Tabel 7. digunakan $k=3$, karena jumlah data yang digunakan pada Tabel 6. sebanyak 5 data.

Tabel 7. Menentukan Bobot *Term*

<i>Term</i>	TF						IDF	Wdt=Tf.idf				
	D1	D2	D3	D4	D5	DF	Log(n/df)	D1	D2	D3	D4	D5
besok	1					1	0,699	0,699	0	0	0	0
bismillah	1					1	0,699	0,699	0	0	0	0
booster	1	1	1	1	1	5	0	0	0	0	0	0
efek			1			1	0,699	0	0	0,699	0	0
gini			1			1	0,699	0	0	0,699	0	0
habis			1			1	0,699	0	0	0,699	0	0
lemes			1			1	0,699	0	0	0,699	0	0
lengan				1		1	0,699	0	0	0,699	0	0
lindung		1				1	0,699	0	0,699	0	0	0
masyarakat		1				1	0,699	0	0,699	0	0	0
mudik		1			1	2	0,398	0	0,398	0	0	0,398
pemerintah		1				1	0,699	0	0,699	0	0	0
pusing			1			1	0,699	0	0	0,699	0	0
syarat		1			1	2	0,398	0	0,398	0	0	0,398
tangan				1		1	0,699	0	0	0,699	0	0

turun				1		1	0,699	0	0	0,699	0	0
udah					1	1	0,699	0	0	0	0,699	0
upaya		1				1	0,699	0	0,699	0	0	0
vaksin	1		1	1	1	4	0,097	0,097	0	0,097	0,097	0,097

Berikutnya adalah tahap menghitung perkalian skalar. Karena yang akan dicari adalah nilai sentimen dari data *tweet* ke lima (D5) maka hitung perkalian skalar antara data *testing* (D5) dengan data *training* (D1,D2,D3,D4) lalu jumlahkan hasil perkalian skalar pada masing-masing data *training* seperti yang ditunjukkan pada Tabel 8.

Tabel 8. Hasil Perkalian Skalar *Data Testing* dan *Data Training*

WD5.WD1			
D1	D2	D3	D4
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0,1584	0	0
0	0	0	0
0	0	0	0
0	0,1584	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0,0094	0	0,0094	0,0094
0,0094	0,3168	0,0094	0,0094

Kemudian langkah berikutnya menghitung kemiripan panjang *vector* setiap *tweet* dengan cara kuadratkan bobot setiap *term* lalu jumlahkan nilai kuadrat tersebut dan diakarkan seperti yang ditunjukkan pada Tabel 9.

Tabel 9. Panjang *Vector* Setiap *Tweet*

Panjang Vector				
D1	D2	D3	D4	D5
0.4886	0	0	0	0
0.4886	0	0	0	0
0	0	0	0	0
0	0	0.4886	0	0
0	0	0.4886	0	0
0	0	0.4886	0	0
0	0	0.4886	0	0
0	0	0	0.4886	0
0	0.4886	0	0	0
0	0.4886	0	0	0
0	0.1584	0	0	0.1584
0	0.4886	0	0	0
0	0	0.4886	0	0
0	0.1584	0	0	0.1584
0	0	0.4886	0	0
0	0	0.4886	0	0
0	0	0	0.4886	0
0	0.4886	0	0	0
0.0094	0	0.0094	0.0094	0.0094
0.9866	2.2712	3.4296	0.9886	0.3262
0.993	1.507	1.851	0.994	0.571

Jika sudah menghitung kemiripan panjang vector dari setiap *tweet*, maka selanjutnya menerapkan persamaan *cosine similarity* sebagai berikut:

	$\text{Cos (D5,D1)} = \frac{0.0094}{0.571 * 0.993} = 0.0166$	(1)
	$\text{Cos (D5,D2)} = \frac{0.3168}{0.571 * 1.507} = 0.3682$	(2)
	$\text{Cos (D5,D3)} = \frac{0.0094}{0.571 * 1.851} = 0.0089$	(3)
	$\text{Cos (D5,D4)} = \frac{0.0094}{0.571 * 0.994} = 0.0166$	(4)

Setelah menerapkan persamaan *cosine similarity*, maka selanjutnya adalah mengurutkan berdasarkan jarak dari nilai tertinggi ke yang terendah seperti yang ditunjukkan pada Tabel 10.

Tabel 10. Jarak Berdasarkan Nilai Terbesar Ke Terkecil

D2	D1	D4	D3
0.3682	0.0166	0.0166	0.0089

Jadi dipilih nilai $k=3$ maka akan dipilih 3 nilai *similarity* yang tertinggi, kemudian ambil sebanyak nilai k ($k=3$) yang paling tinggi tingkat kemiripannya dengan D5 seperti pada Tabel 11.

Tabel 11. *Tweet* Yang Paling Tinggi Kemiripannya

D2	D1	D4
<i>Positive</i>	<i>Positive</i>	<i>Negative</i>

Lalu pilih sentimen yang paling banyak kemunculannya, berdasarkan $k=3$. Sentimen *positive*, diwakili oleh 2 *tweet* yaitu D2 dan D1 sementara sentimen *negative*, diwakili oleh 1 *tweet* yaitu D4.

Dari 3 nilai tertinggi tersebut, kelas sentimen yang paling banyak muncul adalah *positive* sehingga *tweet* D5 masuk ke dalam sentimen *positive*.

3.4 Pengujian

Penelitian ini akan melakukan pengujian dengan menggunakan *confussion matrix* yang diperoleh dari tahapan pemodelan dengan menggunakan algoritma KNN dimana pada tabel 4. akurasi dengan *split data 70:30* lebih tinggi dibandingkan *cross validation*. Sehingga *confussion matrix* yang bisa dilihat pada Tabel 6. adalah hasil dari evaluasi pengukuran klasifikasi *K-Nearest Neighbors* dengan dataset yang berjumlah 2384, terdapat 1669 data training, dan 715 data testing.

Tabel 6. *Confusion Matrix*

	<i>True Positive</i>	<i>True Negative</i>	<i>Class Precision</i>
<i>pred positive</i>	584	99	85.51%
<i>pred negative</i>	7	25	78.12%
<i>class recall</i>	98.82%	20.16%	

Perhitungan dari Tabel 6. adalah sebagai berikut:

	$\text{Precision} = \frac{TP}{TP+FP} = \frac{584}{584+99} = 0.8551 = 85.51$	(1)
	$\text{Recall} = \frac{TP}{TP+FN} = \frac{584}{584+7} = 0.9882 = 98.82\%$	(2)
	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{584+25}{584+25+99+7} = 0.8517 = 85.17\%$	(3)

4. KESIMPULAN

Penelitian ini melakukan ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM VAKSIN *BOOSTER* SEBAGAI SYARAT MUDIK 2022 DENGAN ALGORITMA *K-NEAREST NEIGHBORS*. Berdasarkan hasil pengujian dan analisis yang telah dilakukan, menggunakan data dari media sosial *twitter* pada tanggal 27-30 Mei yang sudah dilakukan *preprocessing* dan pemberian label sentimen dengan jumlah dataset yang digunakan 2384 data dengan sentimen positif 1970 data dan sentimen negatif 414 data. Maka dapat disimpulkan bahwa metode algoritma *K-Nearest Neighbors* mendapatkan nilai akurasi sebesar 85.17% saat menggunakan model pengujian *Split Data* dengan perbandingan 70:30 dan saat model pengujian *Cross Validation* dengan $k=9$ mendapatkan nilai akurasi sebesar 84.4%.

DAFTAR PUSTAKA

- [1] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [2] M. I. Aditama, R. I. Pratama, K. H. U. Wiwaha, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 90–92, 2020.
- [3] R. Rulinawaty, D. Yudhakusuma, P. Hendriarto, Dila Erlianti, and Rini Fitria, "Analisis isi komunikasi publik polri terkait mudik libur natal dan tahun baru," *J. Komun. Prof.*, vol. 6, no. 1, pp. 26–38, 2022, doi: 10.25139/jkp.v6i1.4403.
- [4] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [6] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [7] D. Astuti, "Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/inista.v1i2.71.
- [8] E. Wahyu Sholeha, S. Yunita, R. Hammad, V. Cahya Hardita, T. Rekeyasa Komputer Jaringan, and P. Tanah Laut, "Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naive Bayes dan K-Nearest Neighbor (Sentiment Analysis of Online Travel Agent Using Naive Bayes and K-Nearest Neighbor)," vol. 3, no. 4, pp. 203–208, 2022.
- [9] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
- [10] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.