

Analisis Sentimen pada Aplikasi PeduliLindungi dengan Menggunakan Metode *Improved K-Nearest Neighbor* dan *Lexicon Based*

Nelsih Putriani^{*1)}, Fajri Rakhmat Umbara²⁾, Puspita Nurul Sabrina³⁾

¹⁾²⁾³⁾ Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani
Correspondence author : Nelsih Putriani, nelsihp18@if.unjani.ac.id, Cimahi, Indonesia

DOI: <https://doi.org/10.37012/jtik.v8i1.1107>

Abstrak

Opini atau sentiment masyarakat merupakan salah satu indikator penilaian yang ditujukan untuk menilai suatu hal seperti produk atau jasa. Sentiment tentang aplikasi PeduliLindungi yang banyak digunakan pada masa pandemik ini, perlu dianalisis untuk mengetahui bagaimana opini pengguna aplikasi. Melalui media twitter, sentiment-sentiment masyarakat mengenai aplikasi PeduliLindungi dapat digali, kemudian dideteksi apakah sentiment tersebut termasuk ke dalam sentiment positif, sentiment negative, atau sentiment netral. Salah satu teknik yang digunakan untuk menggali informasi mendeteksi opini masyarakat adalah dengan menggunakan analisis sentiment. Penelitian ini terdiri dari beberapa proses untuk melakukan analisis sentiment, yaitu mengumpulkan data, melakukan tahapan pre-processing, pembobotan kata (ekstraksi fitur), dan proses klasifikasi sentiment ke dalam tiga kelas sentiment. Ketiga kelas tersebut yaitu sentiment positif, negative, dan netral. Dengan menggunakan metode *Improved K-Nearest Neighbor*, penelitian ini memperoleh hasil pengujian analisis sentiment pada tweet berbahasa Indonesia dengan akurasi tertinggi pada k-values 20 sebesar 85%.

Kata kunci : Analisis Sentimen, PeduliLindungi, *Improved K-Nearest Neighbor*, *Lexicon Based*

Abstract

Public opinion or sentiment is one of the assessment indicators aimed at assessing something such as a product or service. Sentiments about the PeduliLindungi application, which is widely used during this pandemic, need to be analyzed to find out the opinions of application users. Through Twitter, people's sentiments regarding the PeduliLindungi application can be explored, then it is detected whether the sentiment includes positive sentiment, negative sentiment, or neutral sentiment. One of the techniques used to gather information to detect public opinion is to use sentiment analysis. This research consists of several processes to carry out sentiment analysis, namely collecting data, carrying out pre-processing stages, weighting words (feature extraction), and the process of classifying sentiment into three sentiment classes. The three classes are positive, negative, and neutral sentiments. Using the Improved K-Nearest Neighbor method, this study obtained sentiment analysis test results on Indonesian-language tweets with the highest accuracy at k-values 20 of 85%.

Keywords: *Sentiment Analysis, Care for Protection, Improved K-Nearest Neighbor, Lexicon Based*

PENDAHULUAN

PeduliLindungi adalah sebuah aplikasi yang dikembangkan untuk membantu instansi pemerintahan dalam melakukan pelacakan yang bertujuan untuk menghentikan penyebaran COVID-19. Aplikasi ini menggunakan partisipasi masyarakat mengenai lokasi terkini atau riwayat lokasi. Hal ini untuk lebih memudahkan dalam melakukan pemantauan

riwayat kontak pengguna dengan penderita COVID-19. Pada aplikasi ini, pengguna akan mendapatkan pemberitahuan mengenai tempat keberadaan pengguna jika sedang berada pada tempat yang rawan atau tempat yang ramai. Area yang rawan tersebut adalah area yang sudah ditandai menjadi zona merah dan terdata bahwa pada area tersebut ada orang yang terinfeksi covid 19 maupun Pasien Dalam Pengawasan. Yang artinya, aplikasi yang digunakan oleh masyarakat ini sudah banyak digunakan pada masa pandemic ini (Aplikasi & Andorid, 2020).

Perkembangan issue dari COVID-19 ramai dibicarakan oleh masyarakat. Terutama dengan adanya aplikasi PeduliLindungi yang erat kaitannya dengan pengontrolan kasus pada masa pandemic COVID-19. Biasanya masyarakat memberikan pendapat maupun opininya melalui media sosialnya (Khusna, 2016). Sehingga pada media sosial akan banyak pandangan mengenai aplikasi ini, apakah aplikasi ini membantu masyarakat dalam masa pandemic atau sebaliknya. Sehingga pendapat yang dikeluarkan bisa dikelompokkan menjadi pendapat yang positif, negative maupun netral (Vadivukarassi, Puviarasan, & Aruna, 2017). Pada media sosial Twitter, PeduliLindungi memiliki akun resmi yang digunakan untuk memberikan informasi terbaru mengenai PeduliLindungi maupun menampung tweet komentar-komentar dari masyarakat (Alsaedi & Khan, 2019). Berdasarkan penelitian PeerReach, Indonesia merupakan negara dengan pengguna Twitter ketiga teraktif di dunia, yang artinya masyarakat Indonesia termasuk pengguna Twitter teraktif di dunia. Jika dilakukan penelitian lebih lanjut pada tweet tersebut maka akan didapatkan sebuah sentimen yang apabila dikumpulkan akan didapatkan sebuah kesimpulan. Analisis sentimen merupakan bidang studi yang melakukan analisis mengenai bagaimana pandangan orang, penilaian, sikap maupun emosi yang dikeluarkan oleh orang tersebut terhadap suatu entitas. Entitas tersebut dapat berupa produk, suatu organisasi, pendapat mengenai individu, masalah yang ada, suatu peristiwa, maupun pandangannya terhadap suatu topik (Srivastava, Singh, & Drall, 2019).

Berdasarkan penelitian (Sodik & Kharisudin, 2021) penggunaan metode k-Nearest Neighbor masih kalah dibandingkan dengan dua metode lainnya. Sehingga penelitian “Implementasi Metode Improved K-Nearest Neighbor Pada Analisis Sentimen Twitter Berbahasa Indonesia (Ramadhan Al-Mubaraq, Al Faraby, & Dwifabri Purbolaksono, 2021)” menggunakan modifikasi K-Nearest Neighbor yaitu *Improved K-Nearest Neighbor* untuk

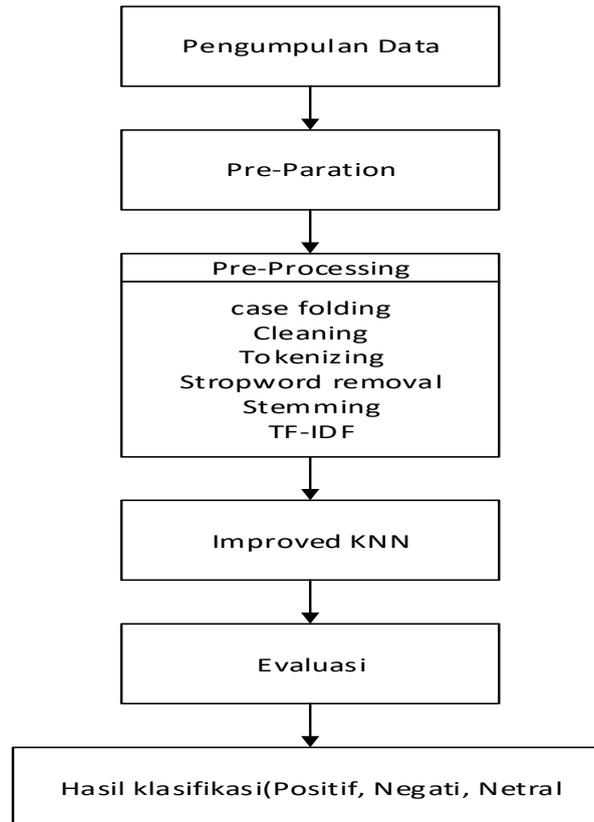
menghasilkan akurasi yang lebih baik. Hasil yang didapatkan melalui penelitian ini adalah, dengan menggunakan metode ini efektivitas sistem yang sudah baik. Karena ada keseimbangan antara data yang digunakan. Serta rata-rata *precision* yang diperoleh sebesar 82%, rata-rata *recall* 87%, dan rata-rata *F-measure* 84%, sehingga dapat disimpulkan bahwa efektivitas sistem sudah berjalan relatif baik. Hal ini dikarenakan penentuan nilai k sangat mempengaruhi nilai akurasi yang dihasilkan (Hermanto & Noviriandini, 2021). Hal inilah yang membedakan diantara keduanya (K-Nearest Neighbor dan Improved K-Nearest Neighbor).

Improved K-NN sudah diimplementasikan pada penelitian mengenai analisis sentiment pada objek penelitian yang berbeda-beda, seperti “Analisis Sentimen Masyarakat terhadap Uji Coba LRT Jakarta Menggunakan Improved K-Nearest Neighbor dan Information Gain” (Putri, Ridok, & Indriati, 2013). Penelitian ini menggunakan kelas sentiment positif dan negative saja serta pelabelan dataset yang digunakan adalah pelabelan secara manual.

Sehingga berdasarkan fenomena tersebut, peneliti tertarik menggunakan metode *Improved K-NN* dan *Lexicon Based* pada analisis sentiment mengenai aplikasi PeduliLindungi dan melihat akurasi yang dihasilkan oleh penggunaan metode ini pada tiga kelas sentiment, yaitu sentiment positif, negative, atau netral. Alasannya pemilihan metode ini yaitu karena melakukan modifikasi k-values untuk menghasilkan akurasi yang lebih stabil. Serta bagaimana mengimplementasikan metode ini untuk objek penelitian yang berbeda, melihat bagaimana keseimbangan data mempengaruhi hasil akurasi yang dihasilkan oleh metode *Improved K-Nearest Neighbor*. Serta berdasarkan penelitian sebelumnya, pelabelan dilakukan dengan cara manual, sehingga peneliti tertarik untuk melakukan pelabelan dataset dengan menggunakan *lexicon based* dan melihat akurasi yang dihasilkan dengan menggunakan metode *Improved K-NN*.

METODE

Berikut merupakan alur metode penelitian yang dilakukan untuk mengetahui opini masyarakat mengenai Aplikasi PeduliLindungi yang secara umum digambarkan pada Gambar 1.



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Data digunakan bersumber dari tweet mention akun @PLindungi dan penggunaan kata “PeduliLindungi”.

2. *Pre-Paration*

Tahap selanjutnya yaitu melakukan persiapan data untuk memasuki tahap *pre-processing* (tahap ini disebut dengan tahap *pre-paration*). Untuk tahapan ini, data disiapkan untuk menentukan kelas sentimennya. Pada tahapan ini penentuan label kelas menggunakan kamus lexicon sebagai pemberian kelas sentiment. Pada pemberian labelling, penggunaan kamus lexicon sudah cukup baik dalam melabelkan suatu sentiment. Hal ini dikarenakan dalam kamus lexicon tersebut sudah memiliki cukup kesesuaian antar kata yang bersifat negative, positif maupun netral.

3. *Pre-Processing*

Tahap *pre-processing* digunakan untuk memilih kata yang nantinya akan digunakan sebagai indeks yaitu kata yang mewakili dokumen. Tahap ini adalah suatu proses pengubahan bentuk kata yang belum terstruktur menjadi kata yang terstruktur. Pada tahap *pre-processing* akan dilakukan proses *case folding*, *cleaning*, *tokenizing*, *filtering*, serta *stemming*.

a. *Case Folding*

Pada tahap *case folding* ini, teks akan diubah dari teks yang menggunakan huruf kapital menjadi teks yang menggunakan huruf lowercase. Serta menghilangkan semua tanda baca yang digunakan (karakter yang diterima hanya huruf “a” sampai huruf “z”, karakter selain huruf tersebut akan diabaikan). Karena tidak semua teks konsisten dalam penggunaan huruf capital (Rohwinasakti, Savira; Irawan, Budi; Setianingsih, 2020).

b. *Cleansing*

Proses *cleansing* bertujuan untuk menghilangkan dokumen yang tidak diperlukan yang bertujuan untuk melakukan pengurangan noise. Kata yang dihilangkan pada proses ini seperti penggunaan hastag (#), username(@), penggunaan emoji, URL (<https://alamat.com>), email dan kata yang tidak diperlukan lainnya.

c. *Tokenizing*

Tokenizing merupakan pemotongan yang dilakukan pada string input berdasarkan setiap kata yang menyusunnya. Pada tahap ini, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan pada kalimat (Mentari, Fauzi, & Muflikhah, 2018).

d. *Filtering / Stopword Removal*

Filtering maupun *Stopword Removal* mempunyai fungsi yang sama. Kedua *pre processing* ini akan membuang kata yang tidak penting dari suatu kalimat. *Stopword removal* dilakukan juga untuk memperbesar akurasi yang nantinya digunakan pada saat pembobotan kata (Pristiyanti, Fauzi, & Muflikhah, 2018).

e. *Stemming*

Stemming merupakan proses pengubahan kata yang menggunakan imbuhan menjadi kata dasar. *Stemming* juga bisa disebut dengan mencari *root* dari setiap kata yang dihasilkan dari proses *filtering*. Pada proses ini kata yang akan diproses

akan dikembalikan pada kata dasarnya, sehingga kata tersebut mampu untuk diolah untuk proses selanjutnya.

f. *Term Frequency – Invers Document Frequency (TF-IDF)*

Bersumber dari Wikipedia.org bahwa *Term Frequency* (TF) merupakan proses pengukuran statistic yang menggambarkan pentingnya istilah tersebut dalam dokumen. Istilah frekuensi sering dibagi dengan panjang dokumen (jumlah total istilah dalam dokumen) sebagai cara normalisasi.

$$TF(w,d) = \frac{\text{jumlah kali istilah } wt \text{ muncul dalam dokumen } t}{\text{jumlah istilah dalam dokumen}} \quad (1)$$

g. *Invers Document Frequency*

Document Frequency Invers (IDF), yang melakukan pengukuran bagaimana pentingnya suatu istilah. Perlu pertimbangan istilah yang sering muncul dan meningkatkan istilah yang langka dengan melakukan perhitungan:

$$IDF(w) = \log \frac{N}{DF} \quad (2)$$

$$TF-IDF(w,d) = TF(w,d) \times IDF(W) \quad (3)$$

Keterangan:

TF-IDF (w,d) bobot suatu kata dalam keseluruhan dokumen

W : suatu kata (word)

d : suatu dokumen

TF(w,d) : frekuensi kemunculan sebuah kata w dalam dokumen

IDF(w) : inverse DF dari kata w

N : jumlah keseluruhan dokumen

DF (w) : jumlah dokumen yang mengandung kata w

4. *Improved K-Nearest Neighbor*

Sebelum mengimplementasikan metode *Improved K-Nearest Neighbor* maka terlebih dahulu diterapkan *cosine similarity* (*CosSim*). Tujuan utama dari *cosine similarity* yaitu untuk menyamakan antara dua dokumen yaitu antara dokumen latih dan dokumen uji (Shanty Wato Wele Kean, 2019). *CosSim* dipandang sebagai pengukuran (similarity

e=measure) antara vector dokumen (D) dengan *vector query*(Q). sehingga, semakin sama dokumen vector dengan *vector query* maka dokumen tersebut dapat dipandang semakin sesuai dengan query. Persamaan yang digunakan untuk menghitung *CosSim* adalah sebagai berikut:

$$\text{cosSim}(x, d_j) = \frac{\sum_{i=1}^m x_i \cdot d_{ji}}{\sqrt{(\sum_{i=1}^m x_i^2)} \times \sqrt{(\sum_{i=1}^m d_{ji}^2)}} \quad (4)$$

Dimana x dokumen dokumen uji, d_j adalah dokumen latih, sementara x_i dan d_{ji} adalah nilai bobot setiap term pada dokumen. Jika nilai yang dihasilkan tinggi maka dokumen uji dengan dokumen latih dikatakan mempunyai kemiripan.

Metode *Improved K-Nearest Neighbor* adalah metode klasifikasi yang digunakan dengan memodifikasi nilai pada *k-values* (Onantya, Indriati, & Adikara, 2019). Penggunaan *improved K-NN* ini mempunyai kelebihan pada jumlah nilai k, sehingga hal inilah yang membuat K-NN dengan Improved K-NN berbeda, yaitu pada penetapan nilai k. Pada K-NN memiliki penetapan nilai k yang sama pada setiap kategorinya. Sedangkan dengan menggunakan Improved K-NN, penetapan nilai k berbeda untuk setiap kategorinya.

Penentuan nilai k yang tepat diperlukan untuk nilai akurasi yang tinggi. Sehingga dengan menggunakan *Improved K-NN* hal yang dimodifikasi dari K-NN sendiri yaitu k-valuesnya. Perbedaan k-values pada setiap kategori disesuaikan dengan seberapa besar kategori yang digunakan pada data latih. Sehingga jika nilai k yang digunakan semakin tinggi, hasil dari pengkategorian tidak terpengaruh pada kategori yang memiliki dokumen dengan jumlah yang maksimal, sehingga pengklasifikasian akan menghasilkan nilai yang lebih baik dengan menggunakan improved K-NN dibandingkan dengan K-NN (Muslimah, Indriati, & Wihandika, 2019).

Perhitungan penetapan k-values pada algoritma Improved K-NN dilakukan dengan menggunakan persamaan:

$$n = \frac{k * N(C_m)}{m * \sum_{j=1}^{N_c} N(C_m)} \quad (5)$$

Dimana n adalah nilai k baru, k adalah k -values yang ditetapkan, $N(C_m)$ adalah jumlah dokumen latih di kategori / kategori m , $\text{Maks}\{N(C_m)|j=1\dots N_c\}$: jumlah dokumen latih terbanyak pada semua kategori.

Dalam menentukan kategori data uji menggunakan *Improved K-NN*, perlu dilakukan perbandingan similaritas pada setiap kategori. Perbandingan ini didasarkan pada proporsi antara dua persamaan di bawah ini:

$$\sum_{d_j \in \text{top } n \text{ kNN}(C_m)} \text{sim}(x, d_j) y(d_j, C_m) \quad (6)$$

$$\sum_{d_j \in \text{top } n \text{ kNN}(C_m)} \text{sim}(x, d_j) \quad (7)$$

Persamaan (6) digunakan untuk menyatakan penjumlahan nilai similaritas sejumlah top n tetangga yang termasuk ke dalam suatu kategori. Sementara persamaan (7) digunakan untuk penjumlahan nilai similaritas top n tetangga pada training set. Selanjutnya data uji akan dilakukan perhitungan $P(x, C_m)$ yang persamaannya adalah:

$$P(x, C_m) = \text{argmax}(m) + \frac{\sum_{d_j \in \text{top } n \text{ kNN}(C_m)} \text{sim}(x, d_j) y(d_j, C_m)}{\sum_{d_j \in \text{top } n \text{ kNN}(C_m)} \text{sim}(x, d_j)} \quad (8)$$

Keterangan:

$P(x, c_m)$: probabilitas dokumen X menjadi anggota kategori c_m

$\text{sim}(x, d_j)$: kemiripan antara dokumen X dengan dokumen latih d_j

top n kNN: top n tetangga

$y(d_j, C_m)$: fungsi atribut dari kategori dengan persamaan

$$y(d_j, C_m) = \begin{cases} 1, & d_j \in C_m \\ 0, & d_j \notin C_m \end{cases} \quad (9)$$

Proses yang dilakukan dalam mengkategorikan data uji menggunakan *Improved KNN* adalah sebagai berikut:

- 1) Langkah pertama yang dilakukan yaitu melakukan proses *preprocessing* agar data siap diolah.
- 2) jika vector sudah terbentuk, selanjutnya melakukan perhitungan *cosine similarity*.

- 3) Melakukan pengurutan nilai dari *cosine similarity*. Pengurutan dilakukan dengan mengurutkan mulai dari nilai terbesar hingga nilai terkecil. Nilai dengan *CosSim* yang besar menandakan kedua dokumen memiliki kemiripan.
 - 4) Selanjutnya jika sudah mendapatkan nilai *CosSim* terbesar maka dilakukan pengelompokan hasil *CosSim* sesuai dengan kategori.
 - 5) Melakukan penentuan k-values baru. Penentuan k-values baru dilakukan juga pada setiap kategori.
 - 6) Setelah top n didapatkan dari langkah sebelumnya, selanjutnya melakukan penentuan kategori data uji.
5. Evaluasi

Dalam penelitian analisis sentiment pada aplikasi PeduliLindungi ini pengujian menggunakan pengujian evaluasi sistem dengan menggunakan akurasi. Hal ini dilakukan untuk mengetahui pengaruh dan banyaknya dokumen latih yang digunakan dan pengaruh k-values pada keberhasilan dalam analisis sentimen.

HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini bersumber dari tweet berbahasa Indonesia yaitu tweet mention yang bersumber dari akun @PLindungi dan penggunaan #pedulilindungi serta penggunaan tweet "PeduliLindungi". Pengumpulan data di dapatkan dengan cara crawling data dari aplikasi RapidMiner. Dengan jumlah data sebanyak 1123 data.

Tahap *pre-paration* digunakan untuk melabelkan suatu data ke dalam label positif, negative, atau netral. Pelabelan ini digunakan dengan *Vader Sentiment* yang memberikan penilaian terhadap data sesuai dengan kata yang ada di dalamnya. Jika nilai polarity sudah digunakan pada data, yaitu $polarity = 0, < 0, > 0$. Maka selanjutnya *value* tersebut dialokasikan pada data *lexicon based* dengan kondisi positif, negative, serta netral. Pada Tabel *pre-paration* di bawah, dapat dilihat bahwa sentiment yang bernilai < 0 artinya sentiment negative, sentiment dengan nilai $= 0$ artinya memiliki sentiment netral. Serta kondisi terakhir yaitu > 0 memiliki sentiment positif.

Tabel 1. Pre-Paration

Tweet	Compound_Score	Sentiment
Padahal aplikasi pedulilindungi bagus untuk tracing	0,4404	Positif
Sistem Satu Data Vaksinasi Covid-19 mencakup integrasi data, sistem keamanan, dan mekansime verifikasi. https://t.co/fUjbKf5WTc	0,34	Positif
Di ingetin resleting kebuka abis nunjukin pedulilindungi ke security. Awkward anyink wkwkwkk	0,6124	Positif
Sangat2 bagus juga nih PeduliLindungi	0,4927	Positif
Pedulilindungi crash mulu	-0,4019	Negatif

Proses pembobotan kata bertujuan untuk memberikan nilai pada setiap kata yang terdapat pada data. Pembobotan ini dilakukan jika proses *pre-processing* sudah dilakukan. Pada pembobotan kata ini, metode yang digunakan yaitu penggunaan TF-IDF. Pada setiap kata yang ada dilakukan pembobotan kata dengan melakukan perhitungan pembobotan kata. Untuk melakukan proses perhitungan bobot kata ini maka dibutuhkan suatu data latih serta data uji, dimana data latih terdapat pada Tabel 1, dan data uji pada Tabel 2.

Tabel 2. Data Latih

Kode	Sebelum Pre-Processing	Setelah Pre-Processing	Sentiment
K1	Padahal aplikasi pedulilindungi bagus untuk tracing	['aplikasi', 'pedulilindungi', 'bagus', 'tracing']	Positif
K2	Sistem Satu Data Vaksinasi Covid-19 mencakup integrasi data, sistem keamanan, dan mekansime verifikasi https://t.co/fUjbKf5WTc	['sistem', 'data', 'vaksinasi', 'covid', 'cakup', 'integrasi', 'data', 'sistem', 'aman', 'mekansime', 'verifikasi']	Positif
K3	Di ingetin resleting kebuka abis nunjukin pedulilindungi ke security. Awkward anyink wkwkwkk	['ingat', 'resleting', 'buka', 'abis', 'nunjukin', 'pedulilindungi', 'security']	Positif
K4	Sangat2 bagus juga nih PeduliLindungi	['bagus', 'pedulilindungi']	Positif
K5	Pedulilindungi crash mulu	['pedulilindungi', 'crash']	Negatif

Tabel 3. Data Uji

Kode	Sebelum Pre-Processing	Setelah Pre-Processing	Sentiment
K6	Error pas dibuka, crash aplikasi pedulilindungi	['error', 'buka', 'crash', 'aplikasi', 'pedulilindungi']	?

Langkah pertama, menghitung jumlah frekuensi yang dimiliki oleh suatu kata pada kalimat.

Tabel 4. Perhitungan *Term Frequency*

Kata	TF					
	K6	K1	K2	K3	K4	K5
Aplikasi	1	1	0	0	0	0
pedulilindungi	1	1	0	1	1	1
Bagus	0	1	0	0	1	0
Tracing	0	1	0	0	0	0
System	0	0	1	0	0	0
Data	0	0	1	0	0	0
Vaksinasi	0	0	1	0	0	0
Covid	0	0	1	0	0	0
Cakup	0	0	1	0	0	0
Integrasi	0	0	1	0	0	0
Aman	0	0	1	0	0	0
Mekanisme	0	0	1	0	0	0
Verifikasi	0	0	1	0	0	0
Ingat	0	0	0	1	0	0
Resleting	0	0	0	1	0	0
Buka	1	0	0	1	0	0
Habis	0	0	0	1	0	0
Tunjuk	0	0	0	1	0	0
Security	0	0	0	1	0	0
Crash	1	0	0	0	0	1
Error	1	0	0	0	0	0

Langkah kedua menghitung TF-IDF

Tabel 5. Perhitungan TF-IDF

Kata	DF	IDF(LOG N/DF)	Wdt = TF.IDF					
			K6	K1	K2	K3	K4	K5
Aplikasi	2	0,477	0,477	0,477	0	0	0	0
pedulilindungi	5	0,079	0,079	0,079	0	0,079	0,079	0,079
Bagus	1	0,778	0	0,778	0	0	0,778	0
Tracing	1	0,778	0	0,778	0	0	0	0
System	1	0,778	0	0	0,778	0	0	0
Data	1	0,778	0	0	0,778	0	0	0
Vaksinasi	1	0,778	0	0	0,778	0	0	0
Covid	1	0,778	0	0	0,778	0	0	0
Cakup	1	0,778	0	0	0,778	0	0	0
Integrasi	1	0,778	0	0	0,778	0	0	0
aman	1	0,778	0	0	0,778	0	0	0
mekanisme	1	0,778	0	0	0,778	0	0	0
verifikasi	1	0,778	0	0	0,778	0	0	0
ingat	1	0,778	0	0	0	0,778	0	0
resleting	1	0,778	0	0	0	0,778	0	0
buka	2	0,477	0,477	0	0	0,477	0	0
habis	1	0,778	0	0	0	0,778	0	0
tunjuk	1	0,778	0	0	0	0,778	0	0
security	1	0,778	0	0	0	0,778	0	0
crash	2	0,477	0,477	0	0	0	0	0
error	1	0,778	0,778	0	0	0	0	0

Untuk dapat mengetahui proporsi data latih pada setiap kategori dan nilai k terhadap ke-efektifan sistem tersebut, maka dilakukan sebanyak lima kali percobaan. Pada penelitian ini, jumlah data uji yang digunakan adalah 310 data, serta penggunaan k-values yang berbeda.

Tabel 6. Hasil Pengujian

K	n(k-value baru)			Akurasi
	Positif	Negatif	Netral	
2	2	1	1	79%
10	10	3	3	82%
20	20	7	6	85%
50	50	17	14	81%
60	60	20	12	63%

Pada tabel di atas dapat dilihat hasil dari pengujian sistem yang sudah dilakukan. Dimana jumlah data yang digunakan pada data latih adalah sebanyak 813 data dan jumlah data uji adalah sebanyak 310 data, dengan jumlah keseluruhan data adalah sebanyak 1.123 data. Pada tabel di atas, akurasi yang paling tinggi di dapatkan dengan k-values sebesar 20 dan jumlah akurasi yang dihasilkan adalah sebesar 85%. Dengan persamaan akurasi yang dapat dilihat sebagai berikut:

$$\text{akurasi} = \frac{\text{jumlah data benar}}{\text{jumlah data keseluruhan}} \times 100\% \quad (10)$$

Dari persamaan tersebut, maka dapat dihitung akurasi yang dihasilkan yaitu sebesar:

$$\text{akurasi} = \frac{264}{310} \times 100\%$$

$$\text{akurasi} = 85\%$$

KESIMPULAN DAN REKOMENDASI

Adapun kesimpulan yang didapatkan dari penelitian ini adalah bahwa penggunaan Metode *Improved K-Nearest Neighbor* adalah metode yang dapat diimplementasikan untuk melakukan klasifikasi pada sentiment termasuk sentiment Bahasa Indonesia. Pengklasifikasian tersebut juga dapat digunakan pada tiga kategori sentiment, yaitu sentiment positif, sentiment negative, dan sentiment netral. Pengujian yang dilakukan pada analisis sentiment aplikasi Pedulilindungi ini menghasilkan akurasi tertinggi dengan *k-values* 20 serta akurasi yang dihasilkan oleh *k-values* tersebut sebesar 85%.

REFERENSI

- Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, 10(2), 361–374. <https://doi.org/10.14569/ijacsa.2019.0100248>
- Aplikasi, L., & Andorid, B. (2020). Mengenal Covid-19 dan Cegah Penyebarannya dengan “Peduli Lindungi ” Aplikasi Berbasis Andorid, (April).
- Hermanto, H., & Noviriandini, A. (2021). Analisa Sentimen Terhadap Belajar Online Pada Masa Covid-19 Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Informatika Kaputama (JIK)*, 5(1), 129–136. Retrieved from <https://jurnal.kaputama.ac.id/index.php/JIK/article/view/451>
- Khusna, I. H. (2016). Opini Publik Cerminan dari Pemerintah dan Kebijakannya. *Promedia*, II(1), 120–136.
- Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(8), 2739–2743.
- Muslimah, N., Indriati, & Wihandika, R. . (2019). Klasifikasi Film Berdasarkan Sinopsis dengan Menggunakan Improved K-Nearest Neighbor (K-NN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(1), 196–204.
- Onantya, I. D., Indriati, & Adikara, P. P. (2019). Analisis Sentimen Pada Ulasan Aplikasi BCA Mobile Menggunakan BM25 Dan Improved K-Nearest Neighbor. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2575–2580. Retrieved from <http://j-ptiik.ub.ac.id>
- Pristiyanti, R. I., Fauzi, M. A., & Muflikhah, L. (2018). Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor, 2(3), 1179–1186.
- Putri, P. A., Ridok, & Indriati. (2013). Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter Berbahasa Indonesia. *Repository Jurnal Mahasiswa PTIHK UB*, 2, 1–8.
- Ramadhan Al-Mubaraq, R., Al Faraby, S., & Dwifebri Purbolaksono, M. (2021). Analisis Sentimen pada Ulasan Film dengan Kombinasi Seleksi Fitur Chi-Square dan TF-IDF

menggunakan Metode KNN, 8(5), 10080–10090.

- Rohwinasakti, Savira; Irawan, Budi; Setianingsih, C. (2020). Analisis Sentimen Pada Produk Layanan Transportasi Online Menggunakan Metode K-Nearest Neighbor, 7(3), 9312–9321.
- Shanty Wato Wele Kean, L. (2019). Analisis Sentimen Review Shopee Berbahasa Indonesia Menggunakan Improved K-Nearest Neighbor dan Jaro Winkler Distance. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(7), 2548–2964. Retrieved from <http://j-ptiik.ub.ac.id>
- Sodik, F., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM , Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. *Prisma*, 4, 628–634.
- Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics*, 14(2), 1–16. <https://doi.org/10.4018/IJHISI.2019040101>
- Vadivukarassi, M., Puviarasan, N., & Aruna, P. (2017). Sentimental Analysis of Tweets Using Naive Bayes Algorithm. *World Applied Sciences Journal*, 35(1), 54–59. <https://doi.org/10.5829/idosi.wasj.2017.54.59>