

Analisis Sentimen Menggunakan *K-Nearest Neighbor* Terhadap *New Normal* Masa *Covid-19* Di Indonesia

Sentiment Analysis using K-Nearest Neighbor towards the New Normal During the Covid-19 Period in Indonesia

Mhd. Furqan¹, Sriani², Susan Mayang Sari³
^{1,2,3}Prodi Ilmu Komputer Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sumatera Utara
E-mail: ¹mfurqan@uinsu.ac.id, ²sriani@uinsu.ac.id, ³susanmys100@gmail.com

Abstrak

New normal diterapkan oleh pemerintah untuk mengembalikan masyarakat beraktivitas normal ditengah pandemi covid-19 dengan protokol kesehatan. Penerapan *new normal* menuai beragam komentar dari masyarakat dan masuk kedalam topik terpopuler di media sosial twitter. Analisis sentimen untuk memprediksi komentar ataupun opini masyarakat yang kecenderungan beropini positif maupun negatif. Preprocessing data menggunakan cleaning, case folding, normalisasi, stemming, filtering, dan tokenizing. Pada normalisasi kata bertujuan memperbaiki kesalahan penulisan kata (typo) berdasarkan KBBI dan TF-IDF sebagai metode pembobotan kata. Data yang digunakan terdiri dari 1000 tweet. Metode klasifikasi opini menggunakan metode *K-Nearest Neighbor* dan melakukan pengujian agar mendapatkan hasil akurasi yang paling terbaik serta mengevaluasi menggunakan confusion matrix. Hasil dari pelabelan untuk sentimen positif berjumlah 811 dan 189 untuk sentimen negatif. Klasifikasi *K-NN* dengan nilai $k = 1$ menghasilkan pengujian use training set dengan accuracy sebesar 100%, 92,60% untuk 10-fold cross-validation dan 94,50% untuk 80% percentage split.

Kata kunci: Sentimen Analisis, *K-Nearest Neighbor (K-NN)*, *New Normal*.

Abstract

The new normal is implemented by the government so that the public can get back to normal activities amid the COVID-19 pandemic with health protocols. The application of the new normal reaped various comments from the society and entered the most popular topics on Twitter social media. Sentiment analysis is to predict review or public opinion that tends to have positive or negative. Preprocessing data using cleaning, case folding, normalization, stemming, filtering, and tokenizing. Word normalization aims to correct typos based on KBBI and TF-IDF as a word weighting method. The data used consists of 1000 tweets. The opinion classification method uses the K-Nearest Neighbor (K-NN) and tests to get the best accuracy results and evaluates using a confusion matrix. The results of labeling for positive sentiment are 811 and 189 for negative. The K-NN classification with a value of $k = 1$ resulted in a use training set with an accuracy of 100%, 92.60% for 10-fold cross-validation, and 94.50% for 80% percentage split.

Keywords: Sentiment Analysis, *K-Nearest Neighbor (K-NN)*, *New Normal*.

1. PENDAHULUAN

Pandemi *covid-19* sudah mewabah virus nyaris ke segala kepenjuru dunia. Tidak hanya mengancam kesehatan serta nyawa, *covid-19* sudah mempengaruhi terhadap segala aspek kehidupan, aspek ekonomi, sosial, pendidikan serta budaya. *World Health Organization (WHO)* melansir bahwa lebih dari 17,6 juta orang terjangkit dan lebih dari 680 ribu orang meninggal dunia di 216 negara. Sedangkan di Indonesia yang terjangkit sudah 130.718 orang, 85.798 orang

yang sembuh dan 5.908 orang yang meninggal dunia [1]. Semenjak kebijakan PSBB diterapkan, banyak menuai reaksi kekecewaan dari masyarakat karena dapat menimbulkan dampak ekonomi yang tidak stabil dan terpaksa melakukan Pemutusan Hubungan Kerja (PHK) sedangkan masih produktif dalam bekerja dan kebijakan ini berpotensi menimbulkan konflik sosial. Penerapan *new normal* yaitu usaha pemerintah untuk mengembalikan masyarakat untuk beraktivitas normal ditengah pandemi *covid-19* dengan menjalankan protokol kesehatan untuk pencegahan wabah tersebut serta upaya pemerintah dalam perihal pemulihan ekonomi. Penerapan *new normal* menuai beragam komentar dari masyarakat, dari sekian banyaknya komentar ada yang beropini positif dan pula negatif serta masuk kedalam daftar topik terpopuler di media sosial twitter. Kondisi tersebut perlunya dilakukan analisis sentimen mengenai *new normal covid-19* pada media sosial twitter dengan mengklasifikasikan secara otomatis opini positif dan negatif untuk berbagai kepentingan serta pengoptimalan informasi pada media sosial sebagai kepentingan publik.

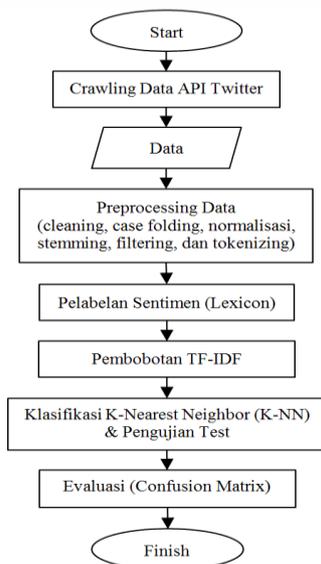
Penelitian terkait pernah dilakukan yaitu analisis sentimen pada siaran tv bersumber pada opini warga di twitter memakai metode *K-Nearest Neighbor*. Penelitian tersebut memperoleh hasil dengan nilai $k=3$ menggunakan pembobotan Tektual yaitu akurasi sebesar 82,50%, 60% untuk Non-Tekstual dan 83,33% untuk gabungan keduanya [2]. Penelitian lainnya yaitu penerapan analisis sentimen twitter yang membahas tentang topik Pilkada DKI 2017 menggunakan metode *K-Nearest Neighbor*. *Preprocessing* yang dilakukan yaitu *case folding*, *tokenizing*, *stopword removal* dan *stemming*. Menggunakan pembobotan kata *TF-IDF* dan fungsi *cosine similarity*. Memperoleh hasil akurasi terbesar 67,2% dengan nilai $k=5$, *precision* tertinggi 56,94% dengan $k=5$ dan *recall* 78,24% dengan $k=15$ [3]. Penelitian selanjutnya yaitu klasifikasi tweets pada twitter menggunakan metode *K-Nearest Neighbor (K-NN)* dengan pembobotan *TF-IDF*. *Preprocessing* yang dilakukan terdiri dari beberapa tahapan yaitu *tokenizing*, *stopword removal* dan *stemming*. Data digunakan berjumlah 140 data dengan uraian 100 data latih dan 40 data uji serta angka nilai k yang dimasukan ialah 1, 3, 5, dan 7. Hasil akurasi dari masing-masing nilai k yaitu $k=1$ akurasi sebesar 75,0%, $k=3$ ketepatan 72,5%, $k=5$ ketepatan 62,5% dan $k=7$ ketepatan 55,0% [4]. Pada penelitian lainnya analisis sentimen pada *review* objek wisata dunia fantasi menggunakan algoritma *K-Nearest Neighbor (K-NN)*. Data diambil dari situs www.tripadvisor.com berjumlah 100 data *review*, terdiri dari 50 *review* positif dan 50 *review* negatif. *Preprocessing* yaitu *tokenize*, *filtering*, *stopword* dan *bi-gram*. Menggunakan pengujian *10-fold cross validation*. Hasil yang diperoleh yaitu akurasi terbesar 77,01 dengan nilai $k=7$, *precision* 92,38%, *recall* 61,56% dan nilai kurva AUC 0.894 [5].

Berdasarkan hasil dari penelitian diatas, maka pada penelitian ini akan melakukan analisis sentimen terhadap *new normal* di era *covid-19* menggunakan *K-Nearest Neighbor (K-NN)* yang merupakan metode sederhana mudah diimplementasikan dan baik dalam mengklasifikasikan sebuah data serta menggunakan metode pembobotan kata *TF-IDF*. Data yang akan digunakan yaitu *tweet* tentang opini masyarakat terhadap kebijakan pemerintah tentang *new normal* yang akan diambil dari media sosial twitter. Data *tweet* akan melalui tahap *preprocessing* yang akan digunakan yaitu *cleaning*, *case folding*, *normalisasi*, *stemming*, *filtering*, dan *tokenizing*. Pada normalisasi kata bertujuan untuk memperbaiki kesalahan pada penulisan kata tidak sesuai (*typo*) berdasarkan KBBI. Untuk mendapatkan nilai akurasi terbaik akan melakukan pengujian tes yaitu *use training set*, *10-fold cross-validation* dan *80% percentage split*. Sebagai evaluasi hasil menggunakan *confusion matriks* yang terdiri dari *accuracy*, *recall*, *precision* dan *f-measure*.

2. METODE PENELITIAN

Prosedur pada klasifikasi dimulai dengan pengumpulan data *crawling api twitter*, persiapan data dengan *preprocessing* data yang terdiri dari beberapa tahapan, yakni *cleaning*, *case folding*, *normalisasi*, *stemming*, *filtering* dan *tokenizing*, proses pelabelan sentimen (Lexicon), proses pembobotan kata *TF-IDF*, proses klasifikasi menggunakan *K-Nearest Neighbor (K-NN)*, dan melakukan pengujian test yaitu *use training set*, *k-fold cross-validation*

dan 80% percentage split. Serta untuk mendapatkan akurasi terbaik menggunakan evaluasi *confusion matrix*. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1 Alur Penelitian

a. Analisis Data

Analisis sentimen ialah pengolahan bahasa natural yang berperan menganalisis pendapat dan juga dapat diartikan sebagai suatu pandangan, keadaan serta perilaku seseorang dalam tulisan ataupun lisan mereka [6]. Analisis sentimen juga dapat sebagai perilaku serta emosi seseorang terhadap suatu produk, penilaian, evaluasi, organisasi, kejadian ataupun topik tertentu [7]. Data itu membagikan informasi untuk bisa membantu seseorang ataupun organisasi dalam pengambilan kesimpulan [8]. Jumlah data yang besar tidak mungkin dianalisis dan dibaca oleh manusia secara manual. Untuk melakukan analisis sentimen dibutuhkan metode yang menunjang klasifikasi, salah satunya yaitu metode *K-Nearest Neighbor (K-NN)*.

Pada penelitian ini penulis menganalisis data sentimen terhadap *new normal* di era *covid-19* dilakukan proses pengumpulan data. Proses pengumpulan data dibatasi hanya mengenai topik ataupun pendapat masyarakat tentang *new normal* saja pada media sosial twitter menggunakan metode *crawling* twitter yaitu memanfaatkan fasilitas *Application Interface (API)* yang tersedia oleh twitter dengan alamat <https://developer.twitter.com>. Data *tweet* yang diambil dari twitter ialah sebanyak 1000 *tweet* dengan keyword "*#NewNormal*", proses pengumpulan data *tweet* hanya mengambil data yang berupa teks berbahasa Indonesia dan tidak mengandung gambar. Pada Gambar 2 diperlihatkan contoh data *tweet* dari proses *crawling* yang digunakan untuk penelitian.

@CommuterLine View 26 September 2020. Di era kenormalan baru, mari utk tidak berbicara secara langsung/melalui telepon genggam saat di dlm KRL. Ayo... #NewNormal <https://t.co/AbdtRIIVg2>

@FaqihHendra3 View 26 September 2020. Negara New Normal, rakyat makin percaya Covid19 itu tidak ada. Pemerintah tuding rakyat tidak disiplin, lha ini wakil rakyat yang bikin. #NewNormal #KhalifahAjaranIslam #Covid19 <https://t.co/BTXjsiA1oo>

@DilaKamilahA15. View 27-09-2020. Tapi juga bisa sebabkan penularan massal plus banyak kematian. Tapi, terima kasih, kita jadi makin yakin isi kepala... <https://t.co/EneBbTNFbr>

Gambar 2 Contoh Data Tweet dari Proses Crawling

b. *Preprocessing Data*

Pada tahapan ini menggunakan data yang berbentuk data teks dari proses *crawling* dan akan melakukan proses *preprocessing* data karena data tersebut masih mempunyai bentuk teks yang tidak terstruktur yang mempunyai banyak *noise*, jadi data itu perlu dibersihkan terlebih dahulu. *Preprocessing* bertujuan untuk meningkatkan kualitas dari data. Jika data itu digunakan untuk proses pencarian pengetahuan kualitasnya rendah, maka pengetahuan yang dihasilkan akan rendah pula [9]. Tahapan *preprocessing* data adalah sebagai berikut :

- 1) *Cleaning* yaitu untuk membersihkan atau menghilangkan karakter yang tidak diperlukan pada data *tweet* seperti tanda baca, *numeric*, *url*, *username*, *mention*, *hashtag*, dan *retweet* seperti (“”~&?!><#%{ }([0-9]+;:;’)[1122].
- 2) *Case Folding* yaitu untuk mengkonversi atau mengubah huruf kapital kedalam huruf kecil (*lowercase*) pada semua data yang terdapat didalam dokumen.
- 3) *Normalisasi* yaitu untuk mengubah dan memperbaiki kata yang disingkat ke dalam kata yang memiliki arti sama berdasarkan KBBI agar menjadi informasi yang dapat diproses dengan mudah misalnya "*utk*" menjadi "*untuk*", "*yg*" menjadi "*yang*" dan sebagainya.
- 4) *Stemming* yaitu untuk menghapuskan seluruh kata imbuhan yang terdapat pada data *tweet* seperti *prefix*, *suffix* dan *konfix* [10].
- 5) *Filtering* yaitu untuk menghapuskan kata yang tidak mengandung makna ataupun kurang penting pada data. Pada tahapan ini menerapkan algoritma *stoplist* atau *stopword*. *Stopword* berisi kata-kata yang biasa dipakai tetapi tidak jelas dan dapat dibuang seperti "*dari*", "*yang*", "*untuk*", "*dan*", "*di*" dan sebagainya.
- 6) *Tokenizing* dilakukan untuk memisahkan *string* atau memecahkan kalimat jadi kata per kata agar mendapatkan kata yang memiliki nilai.

c. *Pelabelan Sentimen*

Pada pelabelan ini data akan diproses secara otomatis yaitu melakukan perhitungan nilai skor dengan kamus *lexicon*. Kamus *lexicon* merupakan metode pengklasifikasian yang menggunakan kamus yang berisi kata opini untuk menentukan suatu sentimen kelas positif dan negatif dari suatu data teks. Kamus pada penelitian ini memakai kamus opini positif dan negatif dari penelitian oleh Liu *et al.*, (2005), yang sudah jadi terjemahan bahasa Indonesia. Menghitung nilai skor pelabelan sentimen dengan metode perhitungan jumlah skor nilai sentimen kelas positif akan dikurangi dengan skor nilai sentimen kelas negatif pada tiap komentar [11]. Proses pelabelan akan dilakukan secara otomatis menggunakan Aplikasi *Rstudio* dan hasil dari perhitungan skor bila sesuatu kalimat yang skornya >0 maka dikategorikan positif, bila kalimat yang skornya <0 maka dikategorikan negatif dan bila kalimat yang skornya =0 maka dikategorikan netral.

d. *Pembobotan Kata TF-IDF*

Pembobotan *TF-IDF* (*Term Frequency-Inverse Document Frequency*) merupakan metode yang melakukan perhitungan pada tiap bobot kata yang terdapat didalam data dokumen. Data dokumen akan diubah menjadi vektor sebanyak kata (*term*) yang akan digunakan untuk klasifikasi. *TF-IDF* digunakan agar bisa melakukan analisis dengan memakai algoritma *K-Nearest Neighbor* (*K-NN*). Adapun perhitungan pembobotan *TF-IDF* ialah sebagai berikut :

$$idf = \log \frac{N}{df} \tag{1}$$

$$w(k, d) = tf(k, d) * idf$$

$$w(k, d) = tf(k, d) * \log \frac{N}{df}$$

Dengan $W(k,d)$ adalah bobot kata (*term*) yang tersedia dalam dokumen, $tf(k,d)$ adalah jumlah frekuensi munculnya kata didalam dokumen, N adalah jumlah seluruh dokumen yang tersedia didalam database dan df adalah jumlah dokumen yang mengandung *term*.

e. *K-Nearest Neighbor (K-NN)*

K-Nearest Neighbor (K-NN) ialah metode sederhana mudah diimplementasikan, data yang digunakan memiliki label sehingga memudahkan dalam proses pengelompokkan ke dalam kelas yang paling sesuai dan memiliki keunggulan dapat mengklasifikasikan data dengan adanya data latih dan data uji serta memiliki kemudahan menerjemahkan hasil dan akurasi dari prediksi dengan secara akurat terlebih dulu nilai *k* terdekat dengan tepat [12]. *K-NN* akan dihitung memakai jarak *Euclidean* dengan persamaan sebagai berikut :

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2} \tag{2}$$

$$= \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Dengan *d(x,y)* sebagai jarak antar variabel *x₁* dan *y₂*, *x* sebagai variabel dan *k* sebagai jumlah variabel.

f. *Evaluasi*

Teknik yang digunakan untuk mengevaluasi hasil yaitu *confusion matrix* banyak digunakan dalam penelitian untuk mengevaluasi hasil dan mengukur kinerja suatu metode klasifikasi dan juga digunakan untuk menghitung serta bisa menarik kesimpulan dari hasil proses penelitian yang sudah dilakukan. Hal ini dibutuhkan untuk menilai seberapa baik sistem dalam proses klasifikasi data. Dalam *confusion matrix* akan dihitung *accuracy*, *precision*, *recall* dan *f-measure* yang dirumuskan pada persamaan (3).

Tabel 1 Confusion Matrix

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \tag{3}$$

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

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

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \times 100\%$$

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan melakukan penerapan algoritma *K-Nearest Neighbor (K-NN)* untuk memperoleh hasil akurasi terbaik dalam mengklasifikasikan data yang berdasarkan pada data sentimen terhadap *new normal covid-19* di media sosial twitter.

a. *Preprocessing Data*

Preprocessing data yang terdiri dari beberapa tahapan, yakni *cleaning*, *case folding*, *normalisasi*, *stemming*, *filtering* dan *tokenizing*. Contoh data hasil proses *preprocessing* data yang dilihat pada Tabel 2.

Tabel 2 Preprocessing Data

Tahapan	Hasil
Data Tweet Asli	Di era kenormalan baru, mari utk tidak berbicara secara langsung/melalui telepon genggam saat di dlm KRL. Ayo... https://t.co/LCdvbxSUPu
Cleaning	Di era kenormalan baru mari utk tidak berbicara secara langsung melalui telepon genggam saat di dlm KRL Ayo
Case Folding	di era kenormalan baru mari utk tidak berbicara secara langsung melalui telepon genggam saat di dlm krl ayo
Normalisasi	di era kenormalan baru mari untuk tidak berbicara secara langsung melalui telepon genggam saat di dalam kereta rel listrik ayo
Stemming	di era normal baru mari untuk tidak bicara cara langsung lalu telepon genggam saat di dalam kereta rel listrik ayo
Filtering	era normal baru mari bicara cara langsung telepon genggam kereta rel listrik ayo
Tokenizing	langsung telepon genggam kereta rel listrik ayo

b. *Pelabelan Sentimen*

Menghitung nilai skor pelabelan sentimen dengan metode perhitungan jumlah skor nilai sentimen kelas positif akan dikurangi dengan skor nilai sentimen kelas negatif pada tiap komentar. Pada penelitian ini terdapat tiga sentimen data yang terdiri dari sentimen positif, negatif dan netral. Namun data yang dipakai hanya sentimen positif dan negatif saja, disebabkan sentimen netral berupa iklan atau spam yang kurang berguna. Contoh hasil perhitungan pelabelan sentimen terhadap *new normal* yang dapat dilihat pada Tabel 3 dan hasil keseluruhan pelabelan data sentimen pada Gambar 3.

Tabel 3 Contoh Perhitungan Pelabelan Kelas Sentimen

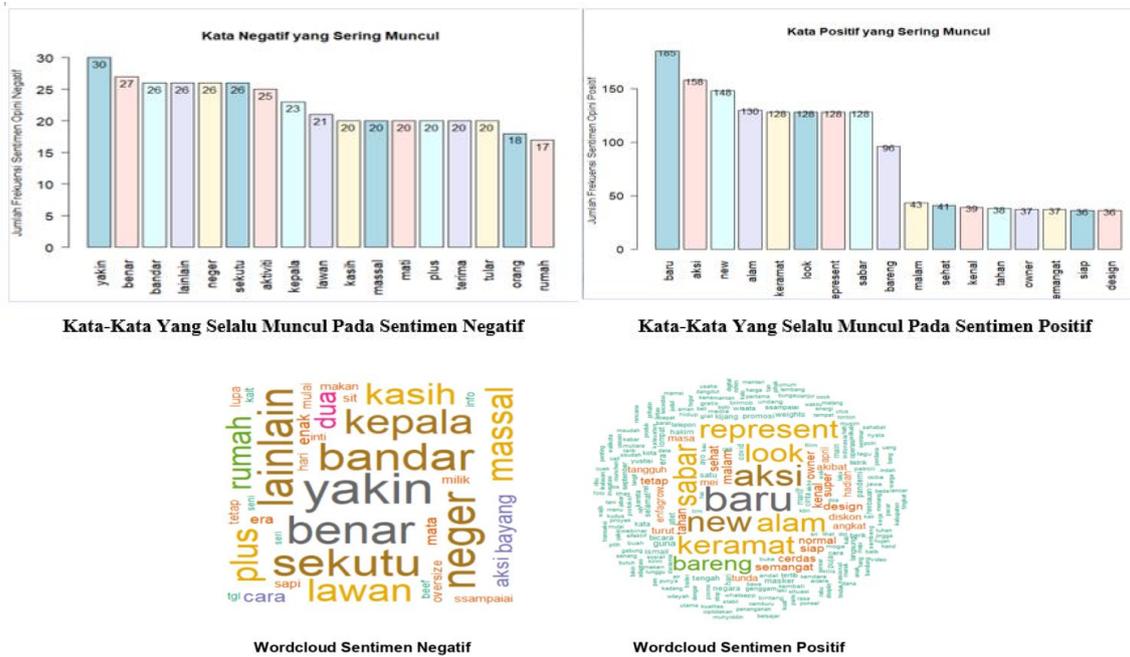
No.	Opini	Kamus Positif	Kamus Negatif	Skor	Kelas Sentimen		
1.	tular massal plus mati terima kasih jadi yakin isi kepala	terima kasih	1	mati tular isi kepala	3	-2	Negatif
2.	era normal baru mari untuk bicara cara langsung telepon genggam kereta rel listrik ayo	normal	1	-	0	1	Positif



Gambar 3 Hasil Kelas Sentimen Terhadap New Normal

Berdasarkan Gambar 3 diatas terlihat jumlah pada data kelas sentimen terhadap *new normal* yang terdiri dari 1000 data opini. Dengan persentase pada kelas sentimen positif mendapatkan 81% dan 19% pada kelas sentimen negatif. Dalam hal ini opini masyarakat terhadap kebijakan *new normal* mendapatkan kelas sentimen positif lebih unggul dibandingkan kelas sentimen negatif. Dengan kata lain masyarakat Indonesia dalam memberi opini atau tanggapan sangat baik terhadap kebijakan *new normal* di era *covid-19*. Hasil pelabelan sentimen terhadap *new normal* dari 1000 data opini pada kelas opini negatif yaitu 189 sentimen dan opini

positif berjumlah 811 lebih unggul dibanding dengan opini bersentimen negatif. Adapun hasil visualisasi untuk keseluruhan data *tweet* opini bersentimen negatif dan positif dapat dilihat pada Gambar 4.



Gambar 4 Visualisasi Pada Sentimen Negatif dan Positif

c. Pembobotan Kata *TF-IDF*

Tahap selanjutnya setelah melewati tahapan pelabelan kelas sentimen adalah tahap pembobotan *TF-IDF*, dimana pada tahapan ini menggunakan teknik perhitungan menggunakan persamaan (1) dimana setiap pembobotan kata (*term*) yang ada didalam data dokumen dihitung dari setiap kata dan setiap kata kemudian akan dikalikan *idf*. Contoh hasil perhitungan *TF-IDF* dapat dilihat pada Tabel 4.

Tabel 4 Contoh Proses Perhitungan *Inverse Document Frequency*

Kata (Term)	TF			DF	N/DF	IDF(log N/DF)
	D1	D2	D3			
tular	1	0	0	1	3	0.477121255
massal	1	0	0	1	3	0.477121255
plus	1	0	0	1	3	0.477121255
mati	1	0	0	1	3	0.477121255
terima	1	0	0	1	3	0.477121255
kasih	1	0	0	1	3	0.477121255
jadi	1	0	0	1	3	0.477121255
yakin	1	0	0	1	3	0.477121255
kepala	1	0	0	1	3	0.477121255

Selanjutnya melakukan tahapan untuk menghitung bobot pada tiap kata. Pada perhitungan bobot (*Weight*) dilakukan perhitungan dokumen ke-*d* terhadap kata (*term*). Sedangkan *tf* merupakan jumlah kemunculan *term* (*t*) dalam dokumen (*d*). *IDF* diperoleh dari hasil perhitungan proses sebelumnya. Hasil dari *term frequency* dikalikan dengan hasil *IDF* akan memperoleh setiap bobot dari masing-masing kata. Contoh perhitungan bobot dapat dilihat pada Tabel 5.

Tabel 5 Contoh Perhitungan Bobot (*Weight*)

W		
D1	D2	D3
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0

d. *Klasifikasi dan Pengujian*

Pada tahap ini akan melakukan perbandingan klasifikasi K-NN dengan $k=1$ terhadap pengujian pada dataset yang telah diperoleh guna untuk mendapatkan hasil akurasi yang terbaik dalam mengklasifikasikan data. Pada pengujian pertama menggunakan *use training set* merupakan pengujian atau pengetesan dengan menggunakan data *training* itu sendiri. Data training akan diproses ke dalam model klasifikasi K-NN kemudian akan melakukan prediksi terhadap 1000 data training. Pengujian *use training set* dalam mengklasifikasikan kelas positif serta kelas negatif telah berhasil diidentifikasi. Kata-kata pada kelas positif terdeteksi dengan benar yaitu sebesar 811 dan 189 kelas negatif. Pada pengujian kedua menggunakan *k-fold cross validation* berguna untuk menghitung kinerja proses suatu algoritma yang membagi dataset jadi beberapa k -buah secara acak serta mengelompokkan data itu sebanyak k -fold. Pada penelitian ini memakai metode *10-fold cross validation*. Cara metode *10-fold cross validation* bekerja yakni membagi data jadi *10-fold* model yang berukuran sama sehingga memiliki 10 subset data yang masing-masing digunakan 9 subset data latih dan 1 subset untuk data uji dengan 10 kali iterasi. Pengujian menunjukkan bahwa hasil sistem klasifikasi K-NN pada pengujian *10-fold cross validation* dengan menggunakan pilihan banyaknya *fold* menghasilkan kelas positif berjumlah 809 dan 117 kelas negatif. Berikutnya pada pengujian ketiga menggunakan *80% percentage split* yang menerapkan pengetesan dengan $k\%$ dari data, dimana k ialah proporsi dari dataset yang digunakan buat data latih. Persentase pada pengujian 80% buat data latih serta 20% buat data uji. Hasil sistem klasifikasi K-NN pada pengujian 80% *percentage split* dalam menerapkan klasifikasi positif berjumlah 172 serta 17 untuk kelas negatif. Perbandingan keseluruhan evaluasi dataset pada sistem pengujian K-NN dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Keseluruhan Evaluasi Pada Pengujian K-NN

Data Set	Algoritma	Mode Test	Correctly Classified Instances		Incorrectly Classified Instances		Mean Absolute Error
			(n)	%	(n)	%	
New Normal	K-Nearest Neighbor (K-NN) Lazy- IBK	Use Training Set	1000	100%	0	0	0,0005
		10 Fold Cross Validation	926	92,6%	74	7,4%	0,0735
		80% Percentage Split	189	94,5%	11	5,5%	0,0539

e. *Evaluasi Hasil*

Evaluasi hasil memakai *confusion matrix* untuk digunakan menghitung nilai *accuracy*, *precision*, *recall* serta *f-measure* menggunakan persamaan (3). Evaluasi hasil bermanfaat buat menyamakan pengujian mana yang cukup baik buat membangun model klasifikasi memakai *K-Nearest Neighbor (K-NN)*.

Tabel 7 Hasil Evaluasi Pengujian Pada *K-NN*

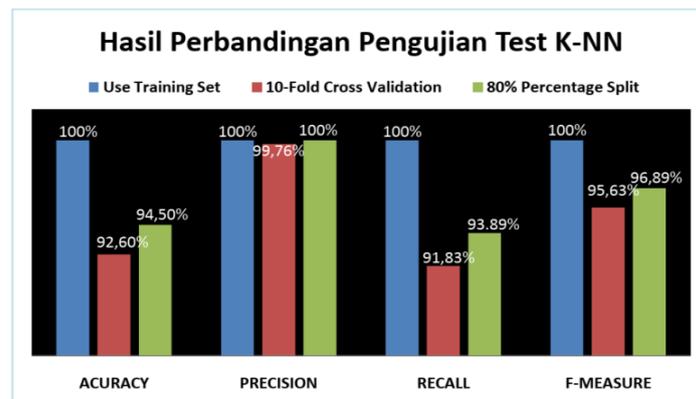
Prediksi	Use Training Set		10-Fold Cross Validation		80% Percentage Split	
	Positif	Negatif	Positif	Negatif	Positif	Negatif
Positif	811	0	809	2	172	0
Negatif	0	189	72	117	11	17

Maka memperoleh nilai evaluasi *confusion matrix* sebagai berikut :

Tabel 8 Hasil Confusion Matrix

Prediksi	Use Training Set	10-Fold Cross Validation	80% Percentage Split
Acuracy	100%	92,6%	94,5%
Precision	100%	99,76%	100%
Recall	100%	91,83%	93,98%
F-Measure	100%	95,63%	96,89%

Setelah mendapatkan hasil perbandingan pengujian *K-NN* bisa dilihat pada diagram di Gambar 5.



Gambar 5. Hasil Perbandingan Pengujian *K-NN*

4. KESIMPULAN DAN SARAN

Analisis sentimen terhadap *new normal* di era *covid-19* pada twitter dengan memakai metode *K-NN* mendapatkan hasil penerapan dalam klasifikasi opini masyarakat menggunakan pelabelan kelas kamus *lexicon* bahwa kelas sentimen positif lebih unggul berjumlah 811 dibandingkan kelas sentimen negatif berjumlah 189. Dengan kata lain masyarakat Indonesia dalam memberi opini atau tanggapan sangat baik terhadap kebijakan *new normal* di era *covid-19*. Dari penelitian ini juga menunjukkan metode *K-Nearest Neighbor (K-NN)* dengan membandingkan pengujian sangat baik dengan nilai $k = 1$. Pada pengujian *use training set* melakukan pengujian menggunakan data *training* itu sendiri mendapatkan hasil akurasi sebesar 100%, *10-fold cross validation* mendapatkan hasil akurasi sebesar 92,60%, dan *80% percentage split* mendapatkan hasil akurasi sebesar 94,50%. Dari hasil penelitian ini penulis menyimpulkan bahwa pengklasifikasian *K-Nearest Neighbor (K-NN)* menggunakan pengujian *80% percentage split* merupakan pengujian yang lebih optimal dibandingkan dengan pengujian lainnya dengan memiliki data latih 80% dan data uji 20% dan dikarenakan juga memiliki nilai akurasi yang tinggi. Saran untuk pengembangan penelitian selanjutnya agar menambahkan pada kamus bahasa Indonesia dan kamus bahasa *slang* dapat ditambahkan untuk penambahan dari kata dasar secara otomatis. Dalam penelitian ini data opini hanya dibatasi untuk opini yang berbahasa Indonesia saja, sehingga perlu dikembangkan kembali menggunakan terjemahan bahasa asing.

DAFTAR PUSTAKA

- [1] Gugus Tugas Percepatan Penanganan Covid-19. <https://covid19.go.id/>. Diakses 18 Agustus 2020. Jam 11:44 WIB.
- [2] Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A., 2017, Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, No.12, Vol.1, 1750–1757, doi: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/631>.
- [3] Deviyanto, A., & Wahyudi, M. D. R., 2018, Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, No.1, Vol.3, 1-13, doi: <https://doi.org/10.14421/jiska.2018.31-01>
- [4] Satrio, R. H., & Fauzi, M. A., 2019. Klasifikasi Tweets Pada Twitter Menggunakan Metode K-Nearest Neighbour (K-NN) Dengan Pembobotan TF-IDF. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, No.3, Vol.8, 8293–8300, doi: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/6133/2950/>
- [5] Sari, R., 2020. Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-NN). *EVOLUSI: Jurnal Sains Dan Manajemen*, No.1, Vol.8, 10–17, doi: <https://doi.org/10.31294/evolusi.v8i1.7371>
- [6] Hartono., Abdullah, D., Hartama, D., Furqan, M., Zarlis, M., & Situmorang, Z., 2016, Sentiment Analysis Using Context Based Fuzzy Linguistic Hedges. *Proceeding International Conference of Computer, Environment, Social Science, Engineering and Technology (ICEST) 2016*, hal 160–162. doi: <http://repository.uinsu.ac.id/8929/>
- [7] Alrumaih, A., Al-Sabbagh, A., Alsabab, R., Kharrufa, H., & Baldwin, J., 2020, Sentiment Analysis Of Comments In Sosial Media. *International Journal Of Electrical And Computer Engineering (IJECE)*, No.6, Vol.10, 5917-5922, doi : <http://ijece.iaescore.com/index.php/IJECE/article/view/22285>.
- [8] Manalu, B.U., 2014, Analisis Sentimen Pada Twitter Menggunakan Text Mining. *Skripsi, Teknologi Informasi Fak. Ilmu Komputer Dan Teknologi Informasi, Universitas Sumatera Utara, Medan*.
- [9] Han, J., & Kember, M., 2006, *Data Mining Concepts and Techniques*. Ed.2. In *Elsevire : San Francisco*.
- [10] Azzahra, A. S., & Wibowo, A., 2020, Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi Dengan Algoritme Naïve Bayes Untuk Ulasan Wisata Kuliner Pada Web Tripadvisor. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, No.4, Vol.7, 737-744, doi: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/1907>.
- [11] Mahendrajaya, R., Buntoro, G. A., & Setyawan, M. B, 2019, Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine. *Komputek : Jurnal Teknik Universitas Muhammadiyah Ponorogo*, No.2, Vol.3, 52–63. doi: <http://studentjournal.umpo.ac.id/index.php/komputek/article/view/270>.
- [12] Jabal Tursina, M., 2019, Sentimen Analisis Sistem Zonasi Sekolah Pada Media Sosial Youtube Menggunakan Metode K-Nearest Neighbor Dengan Algoritma Levenshtein Distance. *Skripsi Fakultas Sains dan Teknologi Program Studi Teknik Informatika, Universitas Islam Negeri Syarif Hidayatullah Jakarta*.
- [13] Nugraha , F. A., Harani, N. H., & Habibi, R., 2020, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif Industri Nusantara : Bandung.