

Implementasi Algoritma *K-Nearest Neighbors* Pada Analisis Sentimen Masyarakat Terhadap Penerapan Pemberlakuan Pembatasan Kegiatan Masyarakat

Vitra Bayu Anwari¹, Yuliazmi^{2*}

^{1,2}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail: ¹vbyuanwari@gmail.com, ^{2*}yuliazmi@budiluhur.ac.id

(* : corresponding author)

Abstrak

Indonesia adalah salah satu negara yang dilanda virus Covid-19, banyak tindakan yang dilakukan pemerintahan dalam memerangi virus Covid-19 seperti menerapkan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) di wilayah DKI Jakarta. Kebijakan tersebut menuai banyak opini masyarakat pada aplikasi *twitter* dan kurang nya informasi yang akurat terkait PPKM yang bisa diperoleh pemerintahan DKI Jakarta dan masyarakat wilayah DKI Jakarta maupun Indonesia, dengan menggunakan data *tweet* dari aplikasi *twitter* akan dilakukan analisis sentimen dari opini masyarakat tersebut dengan menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Total data terkumpul sebanyak 365 data menggunakan aplikasi *rapid miner* dan melewati proses *modeling* menggunakan algoritma *k-nearest neighbors* dengan hasil nilai *accuracy* 83.74%, *precision* 77.78%, dan nilai *recall* 84.00% dari total data sentimen negatif sebanyak 244 dan positif sebanyak 175 data yang berarti analisa yang menggunakan metode *Cross-Industry Standard Process for Data Mining* CRISP-DM dan algoritma *k-nearest neighbors* bisa menghasilkan informasi data yang akurat yang diperuntukkan kepada pemerintahan dan masyarakat wilayah DKI Jakarta terkait penerapan PPKM.

Kata kunci: PPKM, *k-nearest neighbors*, analisis sentimen.

Abstract

Indonesia is one of the countries hit by the Covid-19 virus, many actions have been taken by the government in fighting the Covid-19 virus, such as implementing the Enforcement of Community Activity Restrictions (PPKM) in the DKI Jakarta area. This policy has attracted a lot of public opinion on the twitter application and the lack of accurate information related to PPKM that can be obtained by the DKI Jakarta government and the people of the DKI Jakarta and Indonesia regions, using tweet data from the twitter application, a sentiment analysis of the public opinion will be carried out by applying the Cross methodology. -Industry Standard Process for Data Mining (CRISP-DM). A total of 365 data were collected using the rapid miner application and through the modeling process using the k-nearest neighbors algorithm with the results of 83.74% accuracy, 77.78% precision, and 84.00% recall value of the total 244 negative sentiment data and 175 positive data, which means analysis using the Cross-Industry Standard Process for Data Mining CRISP-DM method and the k-nearest neighbors algorithm can produce accurate data information that is intended for the government and people of the DKI Jakarta area regarding the implementation of PPKM.

Keywords: PPKM, k-nearest neighbors, sentiment analysis.

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang ter dampak virus Covid-19 khususnya wilayah Ibu Kota DKI Jakarta, pemerintah Ibu Kota awalnya membuat kebijakan dengan menerapkan Pembatasan Sosial Berskala Besar (PSBB) yang berdampak pada berbagai sektor kehidupan di Indonesia [1]. Namun seiring dengan berjalannya waktu kebijakan ini dinilai kurang efektif dalam menangani masalah virus Covid-19 yang terus meningkat dan kemudian diterapkan kebijakan lainnya yaitu Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang diterapkan di wilayah zona merah saja, semua pihak berharap kebijakan ini adalah kebijakan yang lebih baik dari sebelumnya dalam menekan penyebaran virus Covid-19 dan sampai saat ini terbukti lebih efektif ketimbang PSBB [2],[3]. Dari berbagai kebijakan yang diterapkan pemerintah DKI Jakarta tentu banyak menimbulkan opini masyarakat yang bernilai positif maupun negatif yang disampaikan melalui pesan *tweet* pada media sosial *twitter* [4].

Dari data *tweet* yang bersisi opini masyarakat tersebut kemudian dikumpulkan untuk dilakukan proses analisis sentimen dengan menerapkan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mempermudah dalam menjalankan tahapan proses *text mining*. Peneliti juga menerapkan algoritma *k-nearest neighbors* untuk pengolahan dan pengelompokan data berdasarkan jarak terdekat antara data *testing* dan data *training* [5].

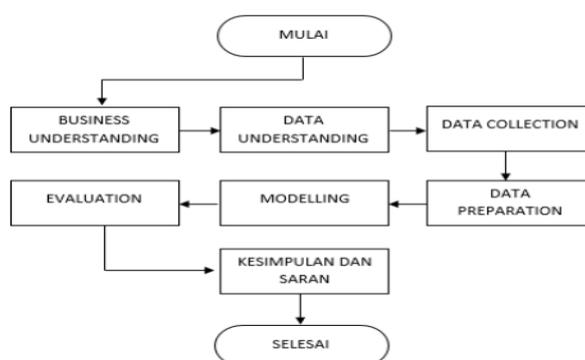
Data yang sudah terkumpul kemudian dilakukan proses *text mining* yang tidak jauh berbeda dengan *data mining* yaitu merupakan proses penambangan data berupa *text* yang akan diolah menjadi informasi yang lebih bermanfaat [6]. Pada penelitian ini juga dibantu oleh seorang pakar ahli pakar dalam memberikan label nilai sentimen dan dilakukan perbandingan dengan algoritma lainnya seperti *naïve bayes* dan *decision tree* untuk mengetahui hasil nilai akurasi terbesar.

Analisis sentimen ini merupakan penelitian yang dilakukan bertujuan untuk mengetahui nilai opini masyarakat dan juga untuk menemukan data berkualitas yang dibutuhkan dan menjadi data yang lebih terstruktur [7],[8]. Data yang dilakukan analisis sentimen merupakan data *tweet* dari sebuah aplikasi *twitter* yang merupakan aplikasi media sosial yang dimiliki oleh Twitter Inc., yang bisa digunakan untuk saling bertukar informasi sesama pengguna dan bisa juga mengirim sebuah foto maupun video di dalam aplikasi tersebut[9]. Tujuan dilakukannya analisis adalah untuk membuat sebuah informasi yang akurat dari sentimen masyarakat terkait penerapan PPKM yang dilakukan pemerintah DKI Jakarta maupun masyarakat DKI Jakarta, Karena permasalahannya adalah tidak adanya informasi yang akurat dan *valid* yang bisa digunakan sebagai data bahan evaluasi dari penerapan PPKM yang dilakukan pemerintahan DKI Jakarta. Sehingga dengan menerapkan metode-metode tersebut bisa mempermudah proses analisa ini dan bisa menghasilkan informasi yang jauh lebih baik dan akurat terkait penerapan PPKM di wilayah DKI Jakarta yang ditujukan untuk pemerintahan wilayah DKI Jakarta sebagai bahan evaluasi maupun info untuk masyarakat DKI Jakarta, dan penelitian ini bisa dijadikan sumber referensi untuk penelitian selanjutnya maupun penelitian lainnya.

Adapun penelitian terdahulu yang dilakukan oleh Moh. Yasid dan Lukman Junaedi [10], penelitian yang berjudul Analisa Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes yang bertujuan untuk mengetahui nilai sentimen masyarakat yang menggunakan jasa Citilink sebagai data bahan evaluasi, menggunakan data dari *twitter* sebanyak 2000 data *tweet* dan terdapat 1661 data yang masuk golongan klasifikasi dan 399 data yang tidak masuk golongan data klasifikasi.

Penelitian selanjutnya yang dilakukan oleh Muhammad Syarifuddin [11], yang berjudul Analisis Sentimen Opini Publik Terhadap Efek PSBB Pada *Twitter* Dengan Algoritma *Decision Tree* – KNN - Naïve Bayes, pada penelitian ini menggunakan tiga algoritma sebagai nilai perbandingan pada penelitian ini juga menggunakan aplikasi *rapid miner* untuk mempermudah memperoleh data pada *twitter*, Hasil nilai tertinggi yang didapatkan adalah hasil dari algoritma *decision tree* dengan nilai *accuracy* 83.3%, *precision* 79%, dan *recall* 81.17% dari total data 170 catatan terkait efek PSBB.

2. METODE PENELITIAN



Gambar 1. Tahapan Metode Penelitian

Berikut adalah tahapan dari metodologi yang diterapkan dalam analisis sentimen yaitu dengan menerapkan metode *text mining* yang merupakan teknik pengembangan data *text* dalam jumlah besar

menjadi pola menarik yang diolah menjadi sebuah informasi yang lebih bermanfaat [12], *text mining* yang tidak jauh berbeda dengan metode *data mining* yang merupakan teknik penggalian informasi dari *big data* baru menjadi data yang lebih bermanfaat dengan menggunakan berbagai teknik, seperti teknik matematika, ilmu statistik, dan *machine learning* [13],[14].

Berikut merupakan tahapan *data mining* yaitu Pembersihan data, Integrasi Data, Seleksi Data, Transformasi Data, Proses Penambangan, dan Presentasi Pengetahuan [15].

2.1. Cross-Industry Standard Process for Data Mining (CRIDP-DM)

Agar tahapan analisis bisa lebih terstruktur peneliti menerapkan *Cross-Industry Standard Process for Data Mining (CRIDP-DM)* yang merupakan metodologi standar industri yang netral dalam proses *data mining* yang dikembangkan pada tahun 1996 oleh DaimlerChrysler, SPSS dan NCR [16], seperti pada Gambar 1. Berikut adalah tahapannya [17] :

a. Pemahaman Bisnis (*Business Understanding*)

Pemahaman bisnis merupakan tahapan awal dimana peneliti melakukan observasi terkait topik penelitian yaitu penerapan PPKM di wilayah DKI Jakarta dan mencari solusi untuk masalah bisnis untuk mencapai tujuan dari penelitian, dan menyusun rencana atau langkah untuk mencapai tujuan penelitian.

b. Pemahaman Data (*Data Understanding*)

Kemudian data yang sudah dipahami secara mendalam, selanjutnya menentukan metode apa yang akan diterapkan dan menentukan *keywords* yang digunakan untuk mengumpulkan data *tweet* pada *twitter*. Seluruh *keywords* yang ditentukan tentu saling berkaitan dengan penerapan PPKM di wilayah DKI Jakarta atau topik penelitian untuk mempermudah dalam proses koleksi data peneliti menggunakan aplikasi *rapid miner* yang merupakan aplikasi *open source* di bawah lisensi GNU *Public License* yang bisa digunakan untuk sistem operasi manapun [18].

c. Koleksi Data (*Data Collection*)

Setelah *keywords* sudah ditentukan, pada tahap ini dimulai tahap pengumpulan data dimulai dari tanggal 8 April 2021 hingga 3 Mei 2021 dengan total data sebanyak 19251 data *tweet* kemudian dilakukan sortir data kembali sehingga data final yang didapatkan sebanyak 365 data *tweet*, kemudian data tersebut akan diberikan oleh pihak ahli pakar untuk keperluan dalam pemberian label nilai sentimen.

d. Persiapan Data (*Data Preparation*)

Setelah data sudah diberikan label nilai sentimen dan di terima kembali oleh peneliti, kemudian data harus melewati *data preprocessing* yang terdiri dari *remove duplicates*, *replace*, *case folding*, *tokenizing*, *stopwords*, *stemming* dan *term weighting* proses ini adalah proses pembersihan data akhir yang dilakukan berkali-kali sampai data benar-benar bersih dan merubah format data sesuai kebutuhan sebelum masuk tahapan pemodelan.

e. Pemodelan (*Modelling*)

Kemudian pada tahap ini adalah tahap dimana peneliti melakukan perhitungan nilai perbandingan antara nilai data yang sudah dilakukan *preprocessing* dan sebelum dilakukan *preprocessing*, yang terdiri dari beberapa tahapan yaitu pemecahan dataset menggunakan *split data* yang terdiri dari data *training* dan data *testing* dengan nilai perbandingan 60:40, 70:30, dan 80:20 dengan menerapkan algoritma *k-nearest neighbors*, *naïve bayes*, dan *decision tree* sebagai nilai perbandingan mana hasil nilai akurasi tertinggi, dan juga menggunakan *cross validation* teknik untuk mendapatkan nilai akurasi terbaik [19].

Kemudian masih pada tahap *modelling* yaitu perhitungan algoritma *k-nearest neighbors* terhadap data yang sudah memiliki nilai sentimen apakah nilai sentimen yang diberikan oleh pakar sesuai dengan perhitungan algoritma yang dilakukan peneliti, algoritma *k-nearest neighbors* diterapkan untuk mengklasifikasikan objek baru berdasarkan jarak terdekat antara data *training* dan

data *testing* dengan syarat nilai *k* harus ganjil dan tidak kurang dari satu [20][21], dengan perhitungan persamaan *cosine similarity* sebagai berikut [22] :

$$\cos(\theta_{ij}) = \frac{\sum_k(d_{ik}d_{jk})}{\sqrt{\sum_k d_{ik}^2} \sqrt{\sum_k d_{jk}^2}} \quad (1)$$

f. Evaluasi (*Evaluation*)

Pada tahapan ini peneliti melakukan evaluasi terhadap hasil perhitungan algoritma *k-nearest neighbors* yang dilakukan terhadap data *tweet* yang sudah melewati proses *data preparation* dengan menggunakan *confusion matrix* yang digunakan untuk menghitung nilai akurasi dari proses klasifikasi apakah sesuai dengan hasil proses *modelling* yang dihasilkan oleh *rapid miner* [23]. Berikut adalah tabel model *confusion matrix* [24] :

Tabel 1. Tabel Model Confusion Matrix

Klasifikasi	+	-
+	TP	FN
-	FP	TN

g. Kesimpulan dan Saran

Kemudian pada tahap ini peneliti mulai menarik kesimpulan yang didapat dari proses dan tahapan dan hasil yang dilakukan pada penelitian ini, adapun saran yang diberikan yang bisa dilakukan penelitian berikutnya.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Tahapan awal ini adalah tahapan dimana peneliti melakukan pemahaman terhadap objek masalah pada penelitian, masalah yang ada adalah tidak adanya informasi yang akurat dan *valid* terkait penerapan PPKM di wilayah DKI Jakarta, maka dari itu penelitian ini akan menerapkan algoritma *k-nearest neighbors* untuk menentukan nilai akurasi terhadap hasil perhitungan nilai akurasi yang merupakan data *tweet* pada aplikasi *twitter* yang saling berkaitan yang memiliki deskripsi penyampaian opini terkait penerapan PPKM di wilayah DKI Jakarta.

3.2. Data Understanding

Kemudian pada tahap ini peneliti mulai melakukan pemahaman terhadap data yang sudah di dapat, karena data yang memiliki deskripsi saja yang bisa dilakukan proses analisis dan bukan dari akun aparat negara, media berita, dan pemerintahan negara, kemudian pada tahap ini peneliti juga melakukan penentuan *keywords* yang akan digunakan untuk proses *crawling* atau pengumpulan data yang saling berkaitan tentang penerapan PPKM di wilayah DKI Jakarta.

3.3. Data Collection

Tahapan ini adalah tahapan di mulainya proses *crawling* data *tweet* pada *twitter* menggunakan aplikasi *rapid miner* dengan *keywords* yang sudah di tentukan pada tahapan *data understanding* yaitu PPKM, PPKM + Mudik, PPKM + Pulang Kampung, PPKM + PSBB, PPKM Mikro, PPKM diperpanjang dengan total data terkumpul sebanyak 19251 data kotor kemudian dilakukan sortir kembali data berkurang menjadi 365 data dalam periode waktu pengumpulan data dimulai dari tanggal 8 April 2021 hingga 20 Mei, data yang dihasilkan pada aplikasi *rapid miner* merupakan data dalam bentuk format file excel seperti gambar berikut :



Gambar 2. Data Hasil *Crawling*

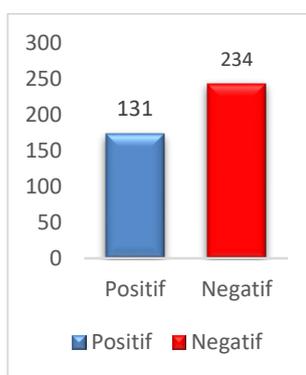
3.4. Data Preparation

- a. Kemudian data hasil *crawling* dilakukan proses *data preprocessing* menggunakan aplikasi *rapid miner*, dimulai dari *remove duplicates, replace, case folding, tokenizing, stopwords, stemming* [25]. Tahapan ini sangat perlu dilakukan selain untuk memfilter dan menyortir untuk mempermudah dalam pemberian label sentimen oleh pakar sebelum masuk tahapan *data preprocessing*, berikut adalah contoh pemberian label sentimen yang ada pada Tabel 2.

Tabel 2. Contoh Data dengan Nilai Sentimen

Tweet	Sentimen
Sukseskan PPKM Mikro	Positif
Diluar rame amat katanya ppkm, hadeuh	Negatif

Berikut adalah gambar persentase sentimen opini masyarakat terhadap penerapan PPKM menurut pakar terhadap data yang diberikan nilai label sentimen



Gambar 3. Persentase Sentimen Masyarakat terhadap Penerapan PPKM

- b. Kemudian Tahapan *data preprocessing* yaitu tahapan dimana peneliti mulai melakukan filter data tahap akhir dan merubah format data berdasarkan format yang dibutuhkan sebelum dilakukan proses perhitungan sebaran data yang sudah dipecah ke dalam data latih dan data percobaan dengan algoritma perbandingan *KNN, naive bayes, dan decision tree* dengan nilai perbandingan 60:40, 70:30, 80:20.

3.5. Modelling

- a. *Modelling* adalah tahapan klasifikasi dengan dua metode yaitu *split data* dengan hasil nilai akurasi tertinggi adalah dengan nilai $k=7$ algoritma *k-nearest neighbors* dan nilai rasio 80:20 sebesar 86.11 % menggunakan data yang sudah dilakukan *preprocessing*, dan metode *cross validation* yang memiliki nilai akurasi 80.76% terhadap data yang sudah dilakukan tahapan *preprocessing*, yang artinya nilai akurasi hasil perhitungan dengan *split data* yang akan dilakukan perhitungan algoritma *k-nearest neighbors* karena memiliki hasil nilai akurasi tertinggi yaitu sebesar 86.11 % dibandingkan hasil nilai akurasi metode *cross validation* sebesar 80.76%
- b. Perhitungan Algoritma *K-Nearest Neighbors*
 Pada tahapan ini peneliti melakukan perhitungan algoritma kembali secara manual dengan contoh 5 data *sample* hasil *crawling* yang memiliki nilai sentimen yang ada pada tabel 3.

Tabel 3. Data Sample

Sample	Tweet	Sentimen
D1	Dukung program PPKM berbasis Mikro	Positive
D2	Dukung dan sukseskan program PPKM Skala Mikro	Positive
D3	Penerapan ppkm hanya mitos pak	Negative
D4	Ppkm mikro tapi wisata, mall buka dongs ??	Negative
D5	PPKM Mikro akan kembali diperpanjang	?

Dengan 5 data *sample* tersebut kemudian adalah menentukan parameter *k* dan menentukan bobot *term* dari 5 data *tweet* yang sudah ditentukan dengan menggunakan *Term Weighting* (TF-IDF) dan nilai *k* = 3 menyesuaikan dengan data *sample* yang berjumlah 5 seperti pada contoh Tabel 4

Tabel 4. Menentukan Bobot Term

Term	K=3						Wdt=tf.idf					
	Tf					Idf	D1	D2	D3	D4	D5	
	D1	D2	D3	D4	D5	Df	Log(n/df)					
dukung	1	1				2	0.398	0.398	0.398	0	0	0
program	1	1				2	0.398	0.398	0.398	0	0	0
sukses		1				1	0.699	0	0.699	0	0	0
ppkm	1	1	1	1	1	5	0	0	0	0	0	0
Basis	1					1	0.699	0.699	0	0	0	0
Penerapan			1			1	0.699	0	0	0.699	0	0
Mikro	1	1		1	1	4	0.097	0.097	0.097	0	0.097	0.097
Skala		1				1	0.699	0	0.699	0	0	0
mitos			1			1	0.699	0	0	0.699	0	0
wisata				1		1	0.699	0	0	0	0.699	0
panjang					1	1	0.699	0	0	0	0	0.699

Tahap selanjutnya adalah perkalian skalar antara data *testing* yaitu D5 dengan data *training* D1, D2, D3, D4, kemudian lakukan penjumlahan hasil perkalian skalar pada data *training* dan data *testing* seperti pada Tabel 5.

Tabel 5. Tabel Perkalian Skalar

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
0.0094	0.0094	0	0.0094
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0.0094	0.0094	0	0.0094

Langkah selanjutnya adalah menghitung nilai kemiripan panjang *vector* dengan cara kuadrat kan bobot setiap *term* lalu jumlahkan nilai kuadrat tersebut kemudian hasil dari penjumlahan nilai kuadrat di akar kan seperti pada tabel 6.

Tabel 6. Panjang Vector

Panjang Vector				
D1	D2	D3	D4	D5
0.158	0.158	0	0	0
0.158	0.158	0	0	0
0	0.488	0	0	0
0	0	0	0	0

0.488	0	0	0	0
0	0	0.488	0	0
0.009	0.009	0	0.009	0.009
0	0.488	0	0	0
0	0	0.488	0	0
0	0	0	0.488	0
0	0	0	0	0.488
0.813	1.301	0.976	0.497	0.497
0.660	1.692	0.952	0.247	0.247

Jika hasil dari perhitungan panjang *vector* sudah ditemukan, maka selanjutnya adalah melakukan perhitungan dengan penerapan *cosine similarity* sebagai berikut:

1. $Cos(D5, D1) = \frac{0.0094}{0.247 * 0.660} = 0.0576$
2. $Cos(D5, D2) = \frac{0.0094}{0.247 * 1.692} = 0.0224$
3. $Cos(D5, D3) = \frac{0}{0.247 * 0.952} = 0$
4. $Cos(D5, D4) = \frac{0.0094}{0.247 * 0.247} = 0.1540$

Setelah melakukan perhitungan dengan menerapkan persamaan *cosine similarity*, maka selanjutnya adalah mengurutkan nilai hasil berdasarkan perhitungan berdasarkan jarak dari nilai hasil terbesar ke hasil yang terendah, seperti yang ditunjukkan pada Tabel 7.

Tabel 7. Jarak Berdasarkan Nilai Terbesar ke Terkecil

D4	D1	D2	D3
0.1540	0.0576	0.0224	0

Kemudian ambil sebanyak tiga nilai dokumen sesuai nilai k ($k=3$) yang paling tinggi tingkat kemiripannya dengan nilai D5, dan eliminasi nilai yang tidak mirip dengan D5, seperti yang ditunjukkan pada Tabel 8.

Tabel 8. Tweet Paling Tinggi Kemiripannya

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

Setelah menentukan nilai yang paling tinggi kemiripannya, kemudian pilih nilai sentimen yang paling banyak kemunculannya berdasarkan nilai $k=3$, seperti yang akan dijelaskan sebagai berikut:

- 1) Dokumen *tweet* D4 merupakan nilai sentimen *negative*
- 2) Sedangkan nilai sentimen *positive* dimiliki oleh 2 Dokumen yaitu D2 dan D1.

Maka kesimpulannya adalah nilai sentimen yang dimiliki D5 adalah *positive* karena jumlah sentimen yang muncul paling banyak pada $k=3$ adalah sentimen *positive*.

3.6. Evaluation

Pada Tahap ini dilakukan pengujian kembali menggunakan *confusion matrix* dari proses tahapan *modelling*, seperti pada Tabel 9.

Tabel 9. *Confusion Matrix*

	<i>True Positive</i>	<i>True Negative</i>	<i>Class Precision</i>
<i>Pred.positive</i>	22	7	75.86%
<i>Pred.negative</i>	8	61	93.02%
<i>Class recall</i>	88.00%	85.11%	

Berikut adalah proses perhitungannya:

1. $Precision = \frac{TP}{TP+FP} = \frac{22}{22+7} = 0.7586 = 75.86\%$
2. $Recall = \frac{TP}{TP+FN} = \frac{22}{22+3} = 0.88 = 88.00\%$
3. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{22+61}{22+61+7+3} = 0.8611 = 86.11\%$

Berdasarkan hasil perhitungan yang dilakukan hasil yang didapatkan *precision* positif adalah 75.86%, *precision* negatif menghasilkan nilai sebesar 93.02%, hasil nilai *recall* sentimen negatif sebesar 85.11%, hasil nilai *recall* positif bernilai 88.00%, jadi kesimpulannya mengapa hasil tidak begitu akurat karena data *training* dan data *testing* yang dimiliki tidak seimbang, dan nilai *accuracy* yang dihasilkan adalah sebesar 86.11%, sehingga dapat disimpulkan berdasarkan hasil perhitungan dan perbandingan bahwa algoritma *k-nearest neighbors* dapat mengklasifikasikan *dataset* sentimen terkait penerapan PPKM di wilayah DKI Jakarta dengan baik.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil yang didapatkan dalam analisis sentimen yang dilakukan dengan menerapkan algoritma *k-nearest neighbors* memiliki nilai akurasi tertinggi yaitu sebesar 86.11% dengan nilai rasio perbandingan 80:20 dan nilai parameter $k=7$ yang sudah dibandingkan dengan hasil perhitungan algoritma *naïve bayes* dan *decision tree* terhadap dataset berupa data *tweet* hasil *crawling* menggunakan *rapid miner* yang memiliki label nilai sentimen yang dilakukan oleh pakar, kemudian hasil dari perhitungan evaluasi menggunakan metode *confusion matrix* memiliki hasil nilai *precision* sebesar 75.86%, *recall* sebesar 88.00%, dan nilai *accuracy* 86.11% terhadap total data sebanyak 365 data dan sebaran data positif sebanyak 131 dan data negatif sebanyak 234 data, Maka dapat disimpulkan berdasarkan hasil yang didapatkan pada penelitian ini dengan menerapkan algoritma *k-nearest neighbors* terkait penerapan PPKM di wilayah DKI Jakarta bahwa hasil yang didapatkan sentimen negatif lebih banyak ketimbang sentimen positif yang berarti penerapan PPKM di wilayah DKI Jakarta masih dianggap belum maksimal berdasarkan opini masyarakat yang disampaikan melalui aplikasi *twitter*, sehingga data tersebut bisa dijadikan bahan evaluasi oleh pemerintah DKI Jakarta dari penerapan PPKM dilakukan, dan juga bisa dijadikan sebuah informasi yang sangat bermanfaat untuk masyarakat DKI Jakarta maupun masyarakat Indonesia.

Adapun saran yang ingin disampaikan terkait dengan penelitian yang dilakukan, yaitu agar melakukan pengembangan dalam menggunakan metode lainnya, dan disarankan agar menggunakan data yang lebih banyak dan lebih update lagi sehingga hasil yang didapatkan bisa lebih optimal dari penelitian sebelumnya. Dan juga perlu dilakukan pengembangan dalam teknik *data preprocessing* sehingga data yang dihasilkan bisa lebih akurat, bersih dan jauh lebih berkualitas karena masih banyak data yang menggunakan Bahasa Asing, Bahasa Daerah, dan Bahasa yang tidak tercatat dalam Kamus Besar Bahasa Indonesia yang tidak terdeteksi oleh sistem.

DAFTAR PUSTAKA

- [1] A. L. Fairuz, R. D. Ramadhani, dan N. A. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial," *Jurnal DINDA*, vol. 1, no. 1, pp. 10–12, 2021.
- [2] P. L. O. Permadhi dan I. M. Sudirga, "Problematika Penerapan Pembatasan Kegiatan Masyarakat (PKM) Kota Denpasar Berbasis Adat Dalam Upaya Penanganan Covid-19," *Prosiding Webinar Nasional Universitas Mahasaraswati Denpasar*, pp.55-60, 2020.
- [3] R. S. Ahmad Gelora Mahardika, "Kedudukan Hukum Pemberlakuan Pembatasan Kegiatan Masyarakat Dalam Sistem Ketatanegaraan Indonesia Ahmad," *LEGACY: Jurnal Hukum dan Perundang-Undangan*, vol. 1, no. 1, pp. 1-23, 2021.
- [4] T. Yulianita, T. W. Utami, dan M. Al Haris, "Analisis Sentimen dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier," *Seminar Nasional Variansi*, pp. 235–243, 2020.
- [5] A. Deviyanto dan M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, pp. 1-13, 2018.
- [6] L. Oktasari, Y. H. Chrisnanto, dan R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Prosiding SNST*, vol. 1, no. 1, pp. 37–42, 2016.
- [7] J. A. Septian, T. M. Fahrudin, dan A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor," *Jurnal of Intelligent System and Computation*, vol.1, no. 1, pp. 43–49, 2019.
- [8] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *Jurnal Teknoinfo*, vol. 14, no. 2, pp. 116-124, 2020.
- [9] R. Rosalina, A. Auzar and H. Hermendra, "Penggunaan Bahasa Slang di Media Sosial Twitter," *Jurnal TUAH*, vol. 2, no. 1, pp. 77–84, 2020.
- [10] M. Yasid, "Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes," *Jurnal Ilmiah Informatika*, vol. 7, no. 2, pp. 82-86, 2019.
- [11] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree-Knn-Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020.
- [12] N. L. Ratniasih, M. Sudarma, dan N. Gunantara, "Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 3, pp. 13, 2017.
- [13] Z. Efendi dan M. Mustakim, "Text Mining Classification sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9, Fakultas Sains dan Teknologi UIN Sultan Syarif Kasim Riau*, pp. 235–242, 2017.
- [14] H. Rofiq, K. C. Pelangi, dan Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes," *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 3, no. 1, pp. 8–15, 2020.
- [15] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques (3rd ed)*. USA: Elsevier, 2011.
- [16] Y. Umidah dan Purwanto, "Penerapan Algoritma K-Nearest Neighbor (KNN) dengan Pencarian Optimal untuk Prediksi Siswa Berprestasi," *JISICOM (Journal of Infomation System, Informatics and Computing)*, vol. 3, no. 2, pp. 1–8, 2019.
- [17] D. T. Larose dan C. D. Larose, *Discovering Knowledge In Data An Introduction to Data Mining*

- Second Edition Wiley Series on Methods and Applications in Data Mining (Second Edition)*. p. 336, 2014.
- [18] S. Haryati, A. Sudarsono, dan E. Suryana, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *Jurnal Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [19] A. Dwiki *et al.*, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [20] I. A. Nikmatun dan I. Waspada, “Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [21] M. Rivki dan A. M. Bachtiar, “Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia,” *Jurnal Sistem Informasi*, vol. 13, no. 1, pp. 31, 2017.
- [22] R. H. Singh, S. Maurya, T. Tripathi, T. Narula, dan G. Srivastav, “Movie Recommendation System using Cosine Similarity and KNN,” *International Journal Engineering. Advanced Technology*, vol. 9, no. 5, pp. 556–559, 2020.
- [23] I. A. Safra and E. Zuliarso, “Analisa Sentimen Persepsi Masyarakat Terhadap Pemandangan Ibukota Baru Di Kalimantan Timur Pada Media Sosial Twitter, Seminar Nasional Multidisiplin Ilmu dan Call For Paper Universitas STIKUBANK (SENDI_U) Ke-6,” pp. 978–979, 2020.
- [24] M. Bramer, *Principles of Data Mining*, Springer, 2007.
- [25] D. Muhidin dan A. Wibowo, “Perbandingan Kinerja Algoritma Support Vector Machine dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 5, no. 2, pp. 153, 2020.