

# Analisis Sentimen terhadap Kebijakan Pemerintah dalam Penanganan Karantina di Indonesia selama Pandemi Menggunakan Algoritma K-Means

1<sup>st</sup> Hafidh Jahfal Prayogo

Universitas Buana Perjuangan Karawang  
Karawang, Indonesia

if18.hafidhprayogo@mhs.ubpkarawang.ac.id  
089664278729

2<sup>nd</sup> Anis Fitri Nur Masruriyah

Universitas Buana Perjuangan Karawang  
Karawang, Indonesia

anis.masruriyah@ubpkarawang.ac.id

3<sup>rd</sup> Elsa Elvira Awal

Universitas Buana Perjuangan Karawang  
Karawang, Indonesia

elsaelvira@ubpkarawang.ac.id

**Abstract**— Penyakit Coronavirus Disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh virus baru yang belum teridentifikasi sebelumnya pada manusia. Terjadinya pandemi virus Corona (COVID-19) di seluruh dunia berdampak pada diberlakukannya karantina bagi setiap orang yang telah melakukan perjalanan lebih dari 2×24 jam. Karantina berfungsi untuk membatasi interaksi antara orang sehat dan yang terpapar COVID-19, serta untuk mengurangi penyebaran COVID-19 yang terjadi di Indonesia. Hasil dari penelitian ini berdasarkan evaluasi Sum of Square Error (SSE), menunjukkan total sebesar 89,7% yang diperoleh dari perhitungan pada tahap TF-IDF dan clustering K-Means. Setelah dilakukan evaluasi Silhouette dengan cluster SSE, diperoleh nilai sebesar 98,3%.

Kata kunci — COVID-19, Karantina, K-Means, SSE, Text Mining

## I. PENDAHULUAN

Terjadinya pandemi virus Corona (COVID-19) di dunia berdampak pada diberlakukannya karantina bagi setiap orang yang telah melakukan perjalanan lebih dari 2×24 jam. Karantina berfungsi untuk membatasi interaksi antara orang sehat dan yang terpapar COVID-19, serta untuk mengurangi penyebaran COVID-19 yang terjadi di Indonesia [1]. Saat ini, Indonesia mengalami pandemi COVID-19, dan pemerintah Indonesia memberlakukan kebijakan terkait karantina. COVID-19 ditetapkan sebagai pandemi oleh Organisasi Kesehatan Dunia (World Health Organization) karena jumlah penyebaran virus bertambah secara signifikan dan berkelanjutan secara global [2]. Selanjutnya, pemerintah Indonesia menetapkan status wabah COVID-19 sebagai Bencana Nasional tentang Corona Virus Disease 2019 (COVID-19) [3].

Media sosial Twitter merupakan bentuk perkembangan teknologi yang menggunakan internet untuk berkomunikasi, saling berbagi informasi, dan membentuk jaringan secara online. Twitter menjadi media sosial yang umum digunakan oleh masyarakat karena perkembangan teknologi yang sangat pesat [4]. Media sosial digunakan sebagai sarana komunikasi sekaligus tempat untuk mendapatkan informasi dan menjadi wadah bagi masyarakat dalam menyampaikan pendapat [5].

Oleh karena itu, pemberlakuan karantina menimbulkan kekhawatiran bagi sebagian besar masyarakat, lantaran banyak informasi yang beredar di media sosial, khususnya Twitter, bahwa pemberlakuan pembatasan kegiatan masyarakat akan diperpanjang menjadi level 3. Maka dari itu, penelitian ini bertujuan untuk menganalisis jumlah respons positif, negatif, atau netral yang disampaikan masyarakat terhadap pemberlakuan pembatasan kegiatan masyarakat melalui media sosial Twitter menggunakan metode algoritma Tidytext [6]. Solusi yang diberikan dalam penelitian ini adalah dengan menyajikan visualisasi data yang diperoleh melalui bahasa pemrograman R atau RStudio.

## II. DATA DAN METODE

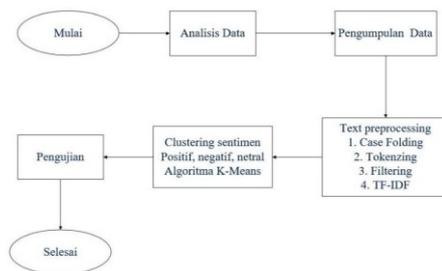
### A. Bahan dan Peralatan

Data digunakan untuk melaksanakan analisis sentimen pada data Twitter bulan Januari 2022 yang memuat informasi kebijakan pemerintah terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Selain itu, terdapat bahan yang digunakan sebagai berikut:

- 1) Perangkat Keras
  - Laptop dengan spesifikasi AMD A4-6210 APU with AMD Radeon R3 Graphics, 4 GB RAM.
- 2) Perangkat Lunak
  - RStudio
  - *Microsoft Office*

### B. Prosedur Penelitian

Tahapan diawali dengan analisis data untuk menerima informasi, lalu dilanjutkan dengan pengumpulan data guna memperoleh data dari Twitter. Selanjutnya, data Twitter akan diproses menggunakan text preprocessing untuk melakukan pengelompokan kata. Setelah itu, data diklasifikasikan menggunakan algoritma K-Means hingga mencapai langkah akhir, yaitu evaluasi prosedur penelitian.



Gambar 1 Prosedur Penelitian

1. Analisis data dilakukan untuk menemukan dan memperoleh sentimen positif, negatif, dan netral.
2. Pengumpulan data diperoleh sejak 4 Januari – 5 Januari 2022 dengan lebih dari 100 data.
3. Dalam tahapan text preprocessing, suatu proses pengubahan bentuk data yang tidak terstruktur menjadi data yang terstruktur sesuai dengan informasi yang dibutuhkan.
4. Algoritma K-Means merupakan salah satu metode pengelompokan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam dua bentuk atau lebih.

C. Akuisisi Data

Data didapatkan dari media sosial Twitter, lalu diakuisisi sebagai data pokok referensi waktu mengadakan proses analisis sentimen dengan jumlah 100 data. Data karantina merupakan tahapan dengan sistem mencari kata yang sering muncul, sentimen positif, negatif, dan netral. Berikut adalah data awal karantina yang telah dimasukkan dalam sistem.

Tabel 1. Data awal karantina

No	KALIMAT
1.	RT @VersaceAlfredo: List ulang: - Cln demam? - Ip Eril"karantina" - Gita cuma bisa SG - Dey masih blm bisa jalan- Olla sakit? - Jesslyn k..."
2.	"RT @LAMPUMERAH_ID: KINAN, ABIS INI LANGSUNG KARANTINAYA..
3.	"@SLAMETSUGIH666 @DWIJAKAANANDA @NXUMXN@ANTTIEEE_ KALO EMNG YG BAWA TAAT KARANTINA DN MPPERHATIKAN BTUL, DIA GAKAN..."
4.	"RT @VERSACEALFREDO: LIST ULANG: - CLN DEMAM? - IP ERIL "KARANTINA" - GITA CUMA BISA SG - DEY MASIH BLM BISA JALAN -OLLA SAKIT? - JESSLYN K..."
5.	"RT @LANGSTONBECKHAM: @BUKEPILING @SUJIWATEDJO BUNAAAAA SEMOGA SUAMIK CEPAT SEMBUH YAA! KALAU SEMUA SEHAT, COBA DEH MINUM WEDHANG UWUH SETIA..."
.	.
96.	NYOKAP GUE KENA COPIT SEKITAR AGUSTUS AWAL, BARU KETAUAN SEMINGGUNYA, PADAHAL GUE KONTAK FISIK TP ALHAMDULILLAH BAN... <a href="https://t.co/2GZMAAXVKH">HTTPS://T.CO/2GZMAAXVKH</a>
97.	MAU JALAN JALAN TAPI GAMAU KENA KONSEKUENSI KARANTINA 10 HARI. FREAK BGT EMG ORG KAYA GITU <a href="https://t.co/fv6evkSF02">HTTPS://T.CO/FV6EVKSF02</a>
98.	RT @VERSACEALFREDO: LIST ULANG: - CLN DEMAM? - IP ERIL "KARANTINA" - GITA CUMA BISA SG - DEY MASIH BLM BISA JALAN -OLLA SAKIT? - JESSLYN K..."
99.	TERPANTAU SUDAH JAM 14.00 BAGIAN RUMAH BABY. BELUM ADA SINOPSIS YG BERTABUR?? SEMOGA BENAR DEH ITU SINOPSIS DIKARAN... <a href="https://t.co/jL9tckQ1IP">HTTPS://T.CO/JL9TCKQ1IP</a>
100.	MAMAS PULANG????TDNYA MAU NGAJAK JALAN2 SOALNYA LG KARANTINA ORG BENERANNYA. TP GABISA JG???? <a href="https://t.co/c2LDIR8F1">HTTPS://T.CO/C2LDIR8F1</a>

D. Text Preprocessing

Tahap text preprocessing berguna untuk mengubah data tekstual yang tidak terstruktur menjadi data terstruktur yang disimpan dalam basis data. Tahapan pada text mining merupakan bagian dari text preprocessing yang bertujuan memperoleh bacaan menjadi data terstruktur. Tahapan yang dilakukan di antaranya:

- 1) Case folding  
Proses mengubah seluruh huruf menjadi huruf kecil.
- 2) Tokenizing

Tokenizing menghilangkan angka, tanda baca, dan karakter lain yang tidak memiliki pengaruh terhadap pemrosesan teks.

3) Filtering

Filtering digunakan untuk membuang kata yang tidak penting dan mengambil kata-kata yang penting.

4) TF-IDF

TF-IDF adalah tahapan terakhir dalam proses text preprocessing untuk melakukan pembobotan kata. Hasil yang didapat dari proses implementasi algoritma menghasilkan data pencilan yang terdapat pada dokumen.

E. Algoritma K-Means

Algoritma metode clustering K-Means merupakan salah satu metode pengelompokan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam dua kelompok atau lebih.

III. HASIL DAN PEMBAHASAN

Penelitian ini memiliki data text preprocessing yang digunakan untuk memperoleh hasil clustering dengan algoritma K-Means, yang menghasilkan sentimen positif, negatif, dan netral.

A. Hasil Text Pre processing

Data diolah menggunakan preprocessing yang diperoleh dari proses implementasi data Twitter ke dalam program, dengan total sebanyak 1.249 data. Tahapan text preprocessing di antaranya to lower, tokenizing, dan filtering. Hasil text preprocessing ditampilkan pada Tabel 2.

No	Proses	Sebelum	Sesudah
1	To lower	"RT @VersaceAlfredo: List ulang: - Cln demam? - Ip Eril "karantina" - Gita cuma bisa SG - Dey masih blm bisa jalan - Olla sakit? -Jesslyn k..."	rt @versacealfredo: list ulang: -cln demam? - ip eril \\karantina\\ - gita cuma bisa sg - dey masih blm bisa jalan - ollasakit? - jesslyn k..."
2	Tokenizing	"RT @VersaceAlfredo: List ulang: - Cln demam? - Ip Eril "karantina" - Gita cuma bisa SG - Dey masih blm bisa jalan - Olla sakit? -Jesslyn k..."	c(";\\rt @versacealfredo: list ulang: - cln demam? - ip eril \\"karantina\\" - gita cuma bisasg - dey masih blm bisa jalan - olla sakit? - jesslyn k...
3	Stemtdocument	"RT @VersaceAlfredo: List ulang: - Cln demam? - Ip Eril "karantina" - Gita cuma bisa SG - Dey masih blm bisa jalan - Olla sakit? -Jesslyn k..."	crt versacealfredo list ulang cln demam ip eril karantina gitacuma bisa sg dey masih blm bisajalan olla sakit jesslyn k...

B. Implementasi Algoritme

Feature Selection adalah sebuah proses untuk menghapus fitur yang berlebihan dan tidak relevan dari data dataset. Tahapan ini memiliki peran yang sangat signifikan dalam akurasi text analytics. Pendekatan yang paling sering digunakan dalam feature selection yaitu Document Frequency (DF), Term Frequency (TF), Inverse Document Frequency (IDF), dan Term Frequency–Inverse Document Frequency (TF-IDF).

1. tidy\_text

Hasil pemrosesan tidy\_text dari RStudio yang ditunjukkan pada Gambar 1.

	nomor	word
1	1	crt
2	1	versacealfredo
3	1	list
4	1	ulang
5	1	cin
6	1	demam
7	1	ip
8	1	eril
9	1	karantina
10	1	gita

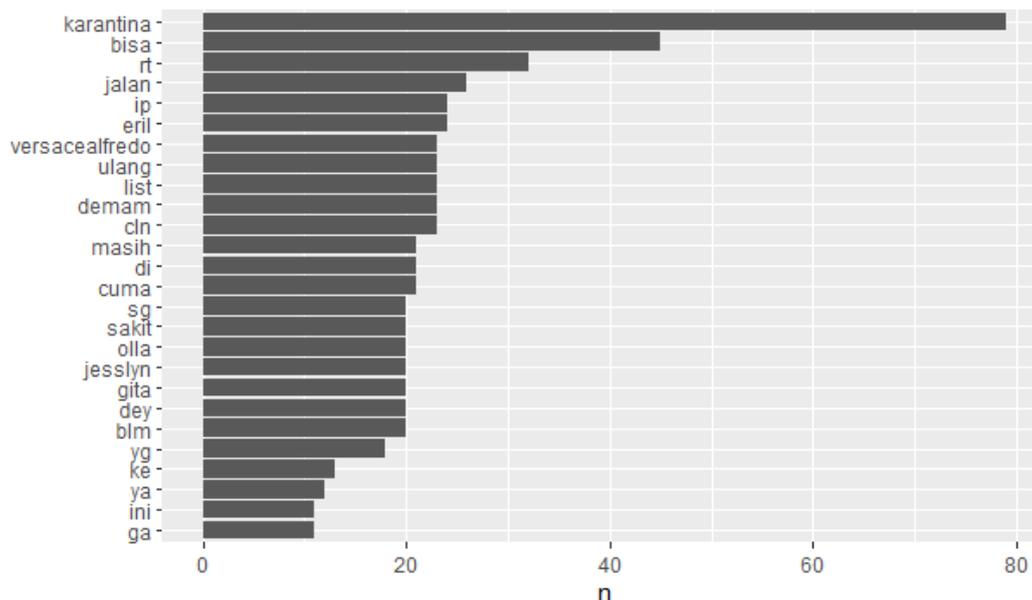
Gambar 1 hasil tidy text

Word cloud tidy\_text terdapat kata “karantina” dengan ukuran terbesar yang menunjukkan kata yang paling sering muncul, serta terdapat kata “bisa” pada Gambar 2.



Gambar 2 word cloud karantina

Dari diagram tidy\_text, kata “karantina” merupakan kata yang paling sering muncul dibandingkan kata lainnya, sedangkan kata “ga” merupakan kata yang paling kecil dan jarang muncul pada Gambar 3.

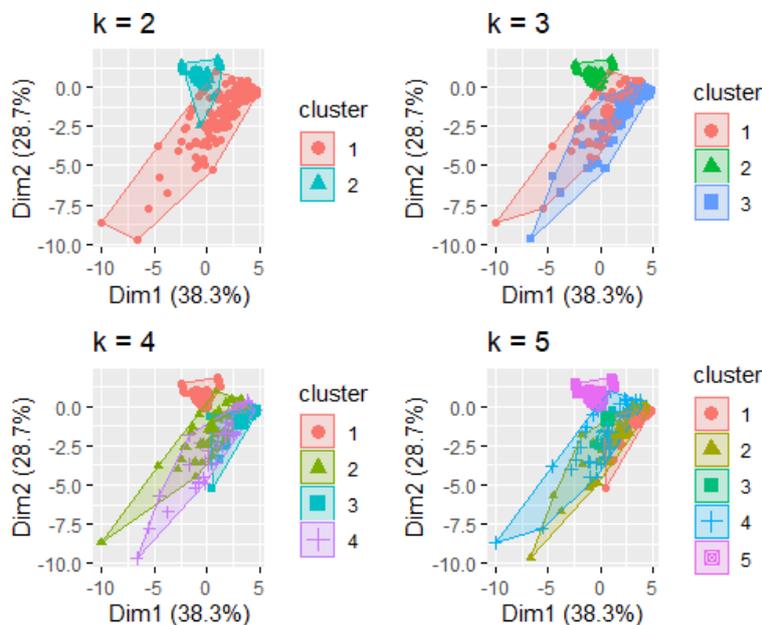


Gambar 3 histogram

Selanjutnya, didapatkan hasil dari klasifikasi dan pengelompokan data dengan masing-masing kata memiliki nilai frekuensi. Terlihat pada Gambar 3, kata-kata yang muncul bersamaan mendekati nilai 80% dari histogram sehingga bersifat positif.

## C. Text Analytic

Proses text analytic pada penelitian ini menggunakan algoritma K-Means untuk mengetahui nilai kluster. Hasil dari analisis yang telah dilakukan menunjukkan jumlah kluster tertentu. Pada  $k = 2$ , kluster 1 lebih besar dibanding kluster 2 karena sentimen netral masuk ke dalam kluster sentimen positif. Selanjutnya, tampak perubahan pada  $k = 3$ , di mana kluster 2 lebih sedikit dibanding kluster 1 dan 3 karena sentimen netral dan positif sudah dibedakan. Kemudian, pada  $k = 4$ , kluster 2 dan kluster 4 sama besar. Pada  $k = 5$ , kluster 4 lebih besar dari kluster lainnya.



Gambar 4 hasil K-Means

Pengujian pada penelitian ini menggunakan hasil dari evaluasi Sum of Square Error (SSE), dengan total 89,7% yang didapat dari perhitungan pada tahap TF-IDF dan clustering K-Means pada Gambar 5.

```
Within cluster sum of squares by cluster:
[1] 4.991623e+29 1.938357e+29
(between_SS / total_SS = 89.7 %)
```

Gambar 5 hasil K-Means

Hasil yang didapat setelah evaluasi dengan cluster SSE terdapat nilai sebesar 98,3%. Nilai tersebut diperoleh dari hasil pengelompokan kata yang terdapat pada Gambar 6.

```
Within cluster sum of squares by cluster:
[1] 3.183352e+27 8.578761e+28 2.478494e+28 6.002692e+06
(between_SS / total_SS = 98.3 %)
```

Gambar 6 evaluasi sse

## IV. KESIMPULAN DAN SARAN

Model yang dihasilkan berbentuk word cloud dan histogram kata yang paling sering muncul, yaitu kata “karantina”. Selanjutnya, model menghasilkan nilai sentimen dari tweet mengenai kebijakan pemerintah terhadap pemberlakuan karantina, yang bersifat positif.

Analisis sentimen memerlukan kamus terbaru untuk mengolah kosakata yang digunakan secara lebih luas, sehingga data yang diproses dalam analisis sentimen menjadi lebih akurat. Penelitian ini terkendala oleh prosesor yang belum mutakhir, sehingga data yang dihimpun tidak lebih dari 100 tweet. Oleh karena itu, pada penelitian selanjutnya disarankan untuk menggunakan teknologi yang lebih mutakhir agar model yang dibangun lebih efektif.

#### PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Hafidh Jahfal Prayogo dengan judul Analisis Sentimen pada Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Algoritma K-Means yang dibimbing oleh Ibu Anis Fitri Nur Masruriyah, M.Kom. dan Ibu Elsa Elvira Awal, M.Kom.

#### DAFTAR PUSTAKA

- [1] Agrani, A., & Rikumahu, B. (2020). Perbandingan Analisis Sentimen Terhadap DigitalPayment ‘Go-Pay’ Dan ‘Ovo’ Di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes Dan Word Cloud Comparison of Sentiment Analysis Against Digital Payment ‘Go-Pay’ and ‘Ovo’ in Social Media Twitter Using N. *Agustus*, 7(2), 2534.
- [2] Alvianda, F., & Adikara, P. P. (2019). Analisis Sentimen Konten Radikal Di Media SosialTwitter Menggunakan Metode Support Vector Machine ( SVM ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 3(1), 241–246.
- [3] Ardiansyah, M. Z. (2019). Pesantren Hybrid Worldview: Moderatisasi Paradigma Penalaran Keislaman dan Pemenangan Kontestasi Wacana Daring. *Jurnal Pendidikan Agama Islam (Journal of Islamic Education Studies)*, 7(1), 1–18. <https://doi.org/10.15642/jpai.2019.7.1.1-18>