

Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan *Naïve Bayes Classifiers*

Taofik Krisdiyanto¹, Erry Maricha Oki Nurharyanto²

^{1,2}Informatika, Universitas Janabadra

Jl. Tentara Rakyat Mataram No.55-57, Bumijo, Kec. Jetis, Kota Yogyakarta,
Daerah Istimewa Yogyakarta 55231

taofik_k@student.janabadra.ac.id¹, errymaricha@janabadra.ac.id²

Abstrak – COVID-19 atau *Coronavirus disease 2019* pada saat ini menjadi sebuah pandemik yang menyebar dengan sangat cepat ke seluruh dunia, termasuk Indonesia. Berbagai penanganan dan kebijakan telah dilakukan salah satunya yaitu kebijakan PPKM atau Pemberlakuan Pembatasan Kegiatan Masyarakat yang dikeluarkan oleh pemerintah Indonesia. PPKM saat ini menjadi salah satu topik yang sering diperbincangkan oleh masyarakat, salah satunya pada *platform* media sosial Twitter. Adanya opini yang diberikan oleh masyarakat, maka perlu adanya analisis sentimen. Analisis sentimen merupakan proses analisis yang diperoleh dari berbagai *platform* media sosial maupun *internet*. Tujuannya yaitu untuk mengetahui bagaimana sentimen masyarakat terhadap penerapan kebijakan PPKM di Indonesia melalui *tweet* dan komentar di *platform* media sosial Twitter. Pada penelitian ini akan dilakukan proses analisis opini masyarakat mengenai kebijakan PPKM dengan mengklasifikasikan opini kedalam 2 sentimen yaitu positif atau negatif. Klasifikasi dilakukan dengan menggunakan metode *Naïve Bayes Classifiers*. Metode *Naïve Bayes Classifiers* merupakan salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan data latih dan data uji yang ada. Hasil penelitian diketahui bahwa opini masyarakat mengenai kebijakan PPKM dihasilkan 99% termasuk kedalam klasifikasi polaritas positif dan 1% polaritas negatif. Selain itu juga dihasilkan bahwa kata yang memiliki frekuensi tertinggi penggunaannya oleh para pengguna twitter adalah kata “darurat”.

Kata Kunci – Analisis Sentimen, Covid19, Kebijakan PPKM, Opini, Media Sosial Twitter, *Naïve Bayes Classifier*.

PENDAHULUAN

Perkembangan teknologi pada saat ini sangat berkembang pesat, salah satu diantaranya

perkembangan *internet*. Perkembangan *internet* yang cepat telah mengubah banyak aspek dalam proses komunikasi data komputer. Dengan hadirnya *internet* sebagai fasilitas yang sangat dibutuhkan membuat semakin banyaknya *platform-platform* yang berkembang dan bisa kita jumpai termasuk *platform* media sosial.

Media sosial merupakan media yang digunakan pengguna dalam rangka mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya. Di dalam media sosial terdapat 3 (tiga) makna bersosial yang diperoleh diantaranya yaitu pengenalan (*cognition*), komunikasi (*communicate*) dan kerja sama (*cooperation*)[1].

Salah satu *platform* media sosial yang populer khususnya di Indonesia yakni Twitter. Twitter merupakan salah satu media sosial yang didirikan oleh Jack Dorsey yang umumnya berfungsi untuk mengirimkan pesan yang disebut dengan kicauan atau (*tweet*)[2]. Penggunaan twitter umumnya diakses para pengguna di Indonesia mencapai 59% dan menduduki peringkat ke-5 media sosial yang sering digunakan pada tahun 2020[3]. Dari data yang ada menjadikan *platform* twitter menjadi media sosial yang cukup memiliki pengaruh bagi para pengguna di Indonesia.

Salah satu topik yang menjadi perbincangan oleh masyarakat saat ini diantaranya di media sosial twitter adalah kebijakan pemerintah mengenai Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). PPKM merupakan kebijakan Pemerintah Indonesia untuk menangani COVID-19 yang dibuat sejak awal tahun 2021[4]. Sebelumnya pemerintah telah melakukan pembatasan sosial berskala besar (PSBB) sebelum adanya PPKM yang dilakukan di sejumlah wilayah Indonesia[5]. PPKM berlangsung di beberapa wilayah yang menjadi titik penyebaran infeksi COVID-19, yakni di Pulau Jawa dan Bali[6].

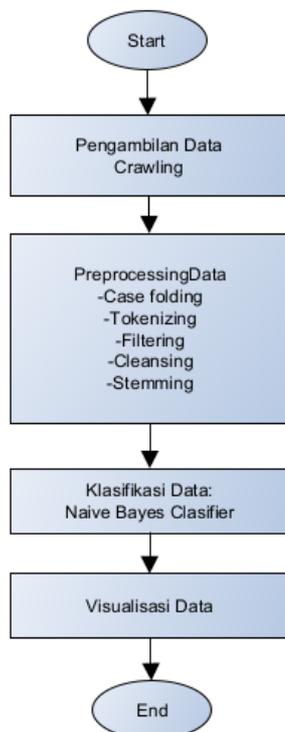
Pemberlakuan PPKM menimbulkan pro-kontra dari masyarakat. Berdasarkan hasil survei SRMC yang diberitakan melalui laman saifulmunjani.com disebutkan bahwa secara

nasional terdapat 44% memilih menjalankan PPKM secara ketat walaupun disisi lain penghasilan menurun, dan terdapat 40% memilih untuk menghentikan PPKM dengan peningkatan risiko penularan COVID-19[7]. Media sosial merupakan media yang memiliki berbagai informasi karena adanya suatu perkembangan teknologi yang memudahkan pengguna untuk mengakses dan digunakan untuk mencurahkan hatinya termasuk permasalahan penerapan PPKM melalui platform tersebut.[8]

Berdasarkan permasalahan yang terjadi, maka menjadi dasar dalam penelitian ini yang bertujuan untuk mengetahui bagaimana sentimen masyarakat terhadap penerapan kebijakan PPKM di Indonesia melalui *tweet* dan komentar di platform media sosial Twitter menggunakan analisis sentimen.

METODE PENELITIAN

Dalam sebuah penelitian terdapat model penelitian yang dibuat dan digunakan sebagai alur perjalanan untuk mempermudah pelaksanaan penelitian. Model penelitian dilakukan berdasarkan beberapa tahapan seperti yang terlihat pada gambar 1.



Gambar 1. Model penelitian

A. Pengumpulan Dataset

Dataset dalam penelitian ini adalah data berbahasa Indonesia mengenai kebijakan PPKM

diperoleh dari twitter dengan hastag "ppkm" dengan metode crawling secara otomatis menggunakan *software* RStudio.

B. Pre-processing

Data yang dikumpulkan merupakan data yang tidak terstruktur sehingga membutuhkan *Pre-Processing*. Proses ini merupakan proses yang sangat penting yang tidak bisa dilewatkan[9] terutama yang berkaitan dengan posts Bahasa daerah yang harus merujuk pada Bahasa Indonesia[10]. *Pre-Processing* dilakukan dalam data melalui beberapa tahapan secara berurutan, sebagai berikut:

1. *Case Folding*
Case Folding merupakan suatu proses untuk mengubah setiap huruf dalam posts menjadi huruf kecil secara keseluruhan atau *lowercase*.
2. *Tokenizing*
Tokenizing merupakan suatu proses pemotongan string input yang dilakukan disetiap kata dalam penyusunannya.
3. *Filtering*
Filtering merupakan proses membuang kata yang tidak memiliki makna atau *stopword*.
4. *Cleansing*
Cleansing merupakan suatu proses untuk menghapus semua karakter dalam posts tidak termasuk alfabet, sehingga akan mengurangi karakter yang tidak diinginkan atau tidak memiliki arti. Karakter tersebut seperti angka, #, @, emoji, maupun link dari suatu situs web yang ada didalam sebuah post.
5. *Stemming*
Stemming merupakan proses mengubah kata yang berimbuhan menjadi kata dasar.

C. Metode *Naïve Bayes*

Naïve Bayes merupakan metode klasifikasi yang berdasar pada teorema Bayes. Metode klasifikasi ini cocok digunakan ketika jumlah masukan yang sangat besar. Klasifikasi ini lebih disukai karena kecepatan dan kesederhanaannya[11]. Dalam klasifikasi ini dikategorikan sebagai klasifikasi sederhana yang sering mencapai suatu performa yang setara dengan algoritma lain seperti *Decision Tree* dan *Neural Network Classifier*. Klasifikasi ini sangat memperhatikan tingginya akurasi serta kecepatan dalam memproses suatu data dalam jumlah yang besar[12].

Klasifikasi *Naïve Bayes* dibangun berdasarkan data latih dalam pembelajaran yang ada dalam algoritma. Tujuan dari algoritma yaitu digunakan untuk memperhatikan perkiraan probabilitas yang ada berdasarkan dari kategori data latih. Klasifikasi *Naïve Bayes* dilakukan dengan menggabungkan pengetahuan yang ada dengan yang baru[13].

Proses klasifikasi *Naïve Bayes* dapat dilihat pada persamaan secara umum[12].

$$P(c_j|w_i) = P(c_j)P_x(wP(iw) | c_j)$$

Keterangan:

- $P(c_j|w_i)$ = Posterior merupakan peluang kategori j ketika terdapat kemunculan kata i
- $P(w_i|c_j)$ = Conditional probability dikategorikan sebagai peluang masuk kata i ke kategori j
- $P(c_j)$ = Prior dikategorikan sebagai peluang munculnya kategori j
- $P(w_i)$ = Peluang kemunculan sebuah kata i
- i = Indeks menjadi suatu kata yang dimulai dari 1 hingga kata k
- j = Indeks kategori yang dimulai dari 1 sampai dengan kategori ke-n

D. Visualisasi

Tahap proses visualisasi ini yaitu memunculkan grafik hasil klasifikasi dengan metode *Naive Bayes Classifier* berdasarkan kategori emosi dan polaritas tiap kata.

HASIL DAN PEMBAHASAN

A. Pengambilan Dataset

Data yang digunakan untuk penelitian ini yaitu data ulasan atau komentar berbahasa Indonesia mengenai kebijakan ppkm dengan hastag “ppkm” pada media sosial yaitu khususnya twitter dengan menggunakan aplikasi RStudio melalui proses *Crawling*. Data yang diambil berjumlah 1000 data ulasan. Hasil data yang didapat saat *crawling* ditunjukkan pada gambar 3.

```
[1] "RT @jokowi: Pemerintah memutuskan untuk memberlakukan PPKM Darurat di Jawa dan Bali, sejak 3 Juli hingga 20 Juli 2021. \n\nPemerintah akan men..."
[2] "Pacitan Ikuti Kebijakan Pusat Terkait PPKM Darurat https://t.co/Do1ENk1r3f"
[3] "Gus Muhaemin mengingatkan soal kebijakan PPKM Darurat yang akan diberlakukan, dia bi..."
[4] "@jokowi PPKM TERUS APA NGKAK ADA SOLUSI LAIN\NPEDAGANG KECIL TAMBAH SUSAH"
[5] "RT @makmumasjid: PSBB\NPSSB TOTAL\NPSSB TRANSISI\NPKKM\NPKKM MIKRO\NPKKM DARURAT\NP..."
[6] "jokowi: Pemerintah memutuskan untuk memberlakukan PPKM Darurat di Jawa dan Bali, sej..."
[7] "RT @budikus11094: Hf gengg, gimana aktifitas kalian? semoga selalu diberikan kelanca..."
[8] "RT @gen987fm: Resmi sudah! \n\nppkm https://t.co/3wduqsrqoP"
```

Gambar 3. Hasil crawling data

Pengambilan *dataset* ini diambil dan disimpan dalam format csv excel. Adapun hasil penyimpanan data berupa csv excel seperti gambar 4 di bawah ini.

```
"*
1,"RT @KawalCOVID19: PPKM Darurat di Jawa dan Bali Bertaku 3-20 Juli, Ini Aturan Lengkapnya | https://t.co/57y1HL7q0 https://t.co/vADANv7Ww"
2,"Penilaian Lomba PPKM Mikro Tingkat Polres Bantul di Kalurahan Poncosari dan Kalurahan Trumurti https://t.co/ALUKRopAW"
3,"RT @satumoneus: PPKM, PPKM Mikro, PPKM Mikro DARURAT... gw list kayak mahasiswa naman file skripsi ni"
4,"RT @indonesiamini: BEKAWA
1. Varian Delta covid-19 yang daya tulanya lebih cepat sudah terdeteksi di banyak tempat.
2. Sekitar 11 Daerah m...
3. "emang gabisa saling salah salah, harus sama sama introspeksi diri, masalahnya di Indonesia ini nah, ga masyarakat... https://t.co/QE3GNkekd"
6,"RT @CNIndonesia: PPKM Darurat Disebut Bikin Gelombang Bomi Gunung Tikar https://t.co/lt0555d0de"
7,"nah baca selengkapnya
PPKM Darurat Demi Rakyat
https://t.co/13gk3r70C"
8,"RT @makmumasjid: pertanyaan soal sejarah UN 2030
1. apa kepanjangan dari PSBB?
2. kapan pertama kali PPKM diterapkan?
3. apa perbedaan PS..."
```

Gambar 4. Hasil penyimpanan data

B. Pre-processing

Tahapan ini merupakan tahap awal untuk mengolah data teks menjadi analisis sentimen. Berikut tahapan *Preprocessing* yang dilakukan:

1. Case folding

Dalam tahapan ini, RStudio akan mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf. Setelah melalui tahap ini, seluruh isi komentar akan menjadi non kapital. Hasil proses *case folding* memperlihatkan susunan kalimat terdiri dari huruf non kapital ditunjukkan oleh gambar 5.

```
> docs[["1"]][["content"]]
[1] "1,RT @naj2709: PPKM DARURAT\ndemi keselamatan bersama, presiden @jokowi akhirnya mem..."
> #case folding
> docs <- tm_map(docs, content_transformer(tolower))
warning message:
In tm_map.simplecorpus(docs, content_transformer(tolower)) :
  transformation drops documents
> view(docs)
> docs[["1"]][["content"]]
[1] "1,rt @naj2709: ppkm darurat\ndemi keselamatan bersama, presiden @jokowi akhirnya mem..."
```

Gambar 5. Hasil proses case folding

2. Tokenizing

Tahap *tokenizing* atau *parsing* merupakan tahap dalam pemotongan *string input* yang dilakukan berdasarkan kata. Proses ini pada prinsipnya digunakan untuk memisah kata yang digunakan dalam penyusunan dokumen. Setiap atau akan terpisah dengan karakter spasi pada umumnya dalam kata lain, sehingga dalam proses *tokenizing* akan mengandalkan karakter spasi yang digunakan untuk proses pemisahan kata[14].

	v1
1	rt ppkm darurat demi selamat sama presiden akhir putu...
2	jika hal ini benar benar terap akan buat ekonomi jakarta st...
3	kutip dari lembar pedoman ppkm darurat yang telah konfi...
4	rt psbb psbb transisi ppkm ppkm mikro penebalan ppk...
5	rt dari psbb gt psbb total gt psbb transisi dari ppkm ...
6	kaburu ppkm f

Gambar 6. Hasil proses tokenizing

3. Filtering

Tahapan setelah proses *tokenizing* adalah *Stopword/stoplist removal*. Hal yang terlebih dahulu dilakukan yaitu me-load file csv yang berisi *stopword* tersebut dan menyimpannya ke dalam sebuah variabel. Pada penelitian ini, file *stopword* berjumlah sebanyak 1045 kata. Data *stopword* ditunjukkan oleh gambar 7.

1398	guoblok
1399	anjay
1400	hrs
1401	ayom
1402	dkk
1403	kopit
1404	zonazi
1405	faksin

Gambar 7. Data stopwords csv

4. Cleansing

Dalam tahapan ini, karakter dalam post secara menyeluruh yang bukan alfabet atau yang tidak memiliki arti sehingga akan berdampak mengurangi adanya karakter dalam analisis sentimen. Hasil proses cleansing ditunjukkan oleh gambar 8.

```
> #menghilangkan data yang kosong dan space di awal dan akhir kalimat
> temp <- as.matrix(docs$content)
> temp <- temp[!apply(temp == " " | temp == "", 1, all),]
> temp <- trimws(temp, which = "both")
> temp <- as.matrix(temp)
> dump <- Corpus(VectorSource(temp))
> view(dump)
> view(temp)
> view(dump)
> dump[["1"]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 52
> dump[["1"]][["content"]]
[1] "darurat presiden putus memberlakukan darurat tanggal"
```

Gambar 8. Hasil proses cleansing

5. Stemming

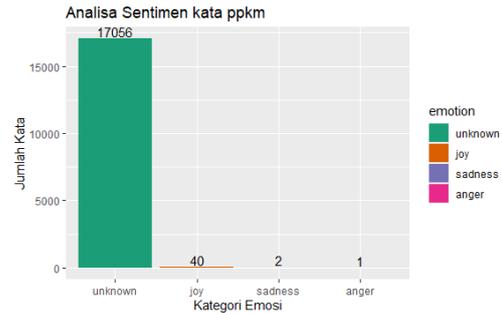
Pada tahap ini, kata dengan kata hubung diubah menjadi kata dasar dengan menggunakan indonesian stemming yang diperoleh dari situs github dengan nama *library* katadasaR ciptaan Nurandi untuk ulasan berbahasa indonesia. Hasil proses *stemming* ditunjukkan oleh gambar 9.

	word	freq
darurat	darurat	809
psbb	psbb	615
mikro	mikro	555
transisi	transisi	251
penebalan	penebalan	208
hidup	hidup	192
jamin	jamin	188
juli	juli	188
pemerintah	pemerintah	123
bal	bal	122
jawa	jawa	119
atur	atur	78
total	total	64
luhut	luhut	64
war	war	64
age	age	61
infinity	infinity	61
ultron	ultron	61
daru	daru	59
putus	putus	55

Gambar 9. Hasil proses stemming

C. Metode Naïve Bayes

Tahap terakhir adalah proses klasifikasi *Naïve Bayes* menggunakan *library* "sentiment". Terdapat dua buah *method* dalam *library* ini yaitu *classify_emotion* dan *classify_polarity*. *Classify_emotion* berfungsi untuk mengklasifikasi kata berdasarkan kategori emosi : "unknown", "joy", "sadness", "fear", "disgust", "anger", dan "surprise". Sedangkan *classify_polarity* berfungsi untuk mengklasifikasi kata berdasarkan kategori polaritas : "positif" atau "negatif". Unknown merupakan hasil dari substitusi untuk menggantikan NA yang berarti tidak memiliki klasifikasi apapun. Hasil proses klasifikasi *Naïve Bayes* ditunjukkan seperti gambar 10.

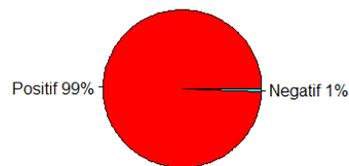


Gambar 10. Hasil proses klasifikasi

D. Visualisasi

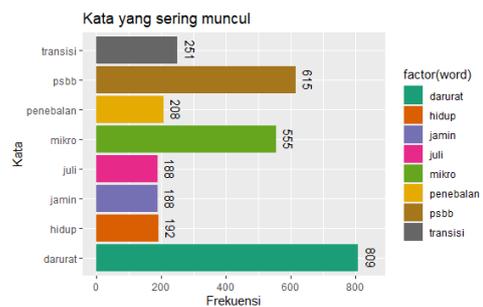
Hasil analisis sentimen yang dilakukan terhadap 1000 dokumen data ulasan dengan metode klasifikasi *Naïve Bayes* Classifier menghasilkan polaritas sentimen positif sebesar 99% dan sentimen negatif 1%. Hasil visualisasi seperti gambar 11.

Analisis Sentimen Kebijakan PPKM



Gambar 11. Hasil analisis sentimen

Berdasarkan hasil *filtering* data menghasilkan 8 kata dengan frekuensi tertinggi yaitu kata "darurat", "hidup", "jamin", "juli", "mikro", "penebalan", "psbb", "transisi". Hasil proses *filtering* frekuensi kata seperti gambar 12.



Gambar 12. Hasil klasifikasi penggunaan kata

Berdasarkan grafik frekuensi penggunaan kata pada gambar 12 disimpulkan bahwa kata yang memiliki frekuensi tertinggi penggunaannya oleh para pengguna twitter adalah kata "darurat" dengan jumlah frekuensi 809 dan hasil tersebut diperkuat dengan hasil yang sama dengan metode penyajian berupa *wordcloud* yaitu kata yang paling menonjol adalah kata "darurat" seperti pada gambar 13.

- [13] K. S. Setyawati *et al.*, “Aplikasi Sentiment Analysis Terhadap Pelaksanaan Pembelajaran Jarak Jauh Universitas Kristen Petra Dengan Metode *Naïve Bayes Classifier*.”
- [14] F. Nurhuda, S. Widya Sihwi, and A. Doewes, “Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode *Naïve Bayes Classifier*,” *J. Teknol. Inf. ITSmart*, vol. 2, no. 2, p. 35, 2016, doi: 10.20961/its.v2i2.630.