

# IMPLEMENTASI METODE LEXICON BASE UNTUK ANALISIS SENTIMEN KEBIJAKAN PEMERINTAH DALAM PENCEGAHAN PENYEBARAN VIRUS CORONA COVID-19 PADA TWITTER

Abdus Syakur

Fakultas Teknologi Industri Universitas Gunadarma  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat  
syakur@staff.gunadarma.ac.id

## Abstrak

*Penyebaran virus corona COVID-19 semakin meluas ke seluruh dunia, termasuk negara Indonesia. Upaya pemerintah dalam menangani pandemi ini salah satunya adalah dengan menerapkan kebijakan-kebijakan terkait penyebaran virus. Masyarakat Indonesia yang majemuk mempunyai opini yang beragam terhadap pemberlakuan kebijakan tersebut. Analisis sentimen merupakan studi yang menganalisis opini, sentimen dan emosi seseorang terhadap sebuah produk atau topik yang bertujuan untuk mendapatkan informasi opini masyarakat dan kemudian mengklasifikasikannya. Tujuan dari penelitian ini adalah menemukan informasi opini masyarakat terhadap kebijakan pemerintah dalam menghadapi pandemi COVID-19 dengan harapan pemerintah dapat memantau dan mempertimbangkan kebijakan yang diberlakukan. Penelitian ini menerapkan metode lexicon-based untuk menganalisis polaritas opini masyarakat. Pengumpulan data dilakukan dari tanggal 17 April hingga 24 April 2020 dengan hashtag jokowi. Dalam penelitian ini penulis memanfaatkan library textblob yang menyediakan kamus berisi leksikon sentimen. Tahap preprocessing yang dilakukan adalah case folding, menghapus string URL, menghapus karakter newline, menghapus mention dan retweet, mengganti nama pengguna dengan spasi, menghapus hashtag, replace slang, penghapusan tanda baca dan menerjemahkan teks kedalam Bahasa Inggris. Berdasarkan hasil analisis dihasilkan informasi bahwa sebagian besar masyarakat Indonesia memberikan tanggapan positif terhadap kebijakan pemerintah dalam upaya pencegahan penyebaran virus corona COVID-19 di Indonesia. Persentase menunjukkan 30,7% tanggapan positif, 14,3% tanggapan negatif, dan 55% netral.*

**Kata Kunci:** analisis sentimen, lexicon based, covid-19, twitter, kebijakan pemerintah

## Abstract

*The spread of the COVID-19 corona virus is widespread throughout the world, including the country of Indonesia. One of the government's efforts to deal with the pandemic is to implement policies related to the spread of the virus. The diverse Indonesian society has diverse opinions on the implementation of the policy. Sentiment analysis is a study that analyzes a person's opinions, sentiments and emotions on a product or topic that aims to obtain information on public opinion and then classifies it. The purpose of this study is to find public opinion on government policies in dealing with the COVID-19 pandemic, hopefully the government can monitor and consider the policies that are implemented. This study applies a lexicon-based method to analyze the polarity of public opinion. Data collection was carried out from April 17 to April 24, 2020 with hashtag jokowi. In this study the authors utilize the textblob library which provides a dictionary containing the sentiment lexicon. The preprocessing steps include case folding, deleting the URL string, deleting newline characters, deleting mention and retweeting, replacing user names with spaces, deleting hashtags, replacing slang, deleting punctuation and translating text into English. Based on the results of the analysis produced information that most of the Indonesian people gave a positive response to government policies in efforts to prevent the*

*spread of the COVID-19 corona virus in Indonesia. The percentage shows 30,7% positive responses, 14,3% negative responses, and 55% neutral.*

**Keywords:** *sentiment analysis, lexicon based, covid-19, twitter, government policy*

## PENDAHULUAN

Informasi mengenai penyebaran virus corona banyak ditemukan di media digital maupun media cetak konvensional. Di Indonesia, media sosial Twitter masih terbilang cukup banyak digunakan. Menurut hasil riset *Wearesosial Hootsuite* yang dirilis Januari 2019, Twitter menempati urutan empat setelah Youtube, Facebook dan Instagram sebagai media sosial yang banyak digunakan di Indonesia, hasil riset mencapai 52% dari total populasi. Menurut Latifa [1] “Pengguna Twitter yang berasal dari berbagai kalangan dan disiplin ilmu menyebabkan aliran komunikasi dan opini di media ini menjadi lebih beragam, salah satunya adalah kritikan dan komentar mengenai kebijakan pemerintah”. Kebijakan perintah terkait upaya pencegahan penyebaran virus corona banyak menarik perhatian masyarakat. Kebijakan tersebut bahkan banyak menimbulkan polemik ditengah masyarakat. Pemerintah tentu akan memantau dan mempertimbangkan pemberlakuan kebijakan berdasarkan respon dari masyarakat. Masyarakat mengungkapkan melalui Twitter yang mengakibatkan banyak sekali respon terkait kebijakan perintah dalam upaya pencegahan penyebaran virus corona ini. Dalam setiap *tweet* dapat diklasifikasikan menjadi dua jenis *tweet* yaitu *tweet* dengan

sentimen positif dan *tweet* dengan sentimen negatif.

Analisis sentimen merupakan analisis untuk mengidentifikasi apa yang masyarakat suka, tidak suka, pendapat, komentar atau umpan balik yang di klasifikasikan menjadi respon positif, negatif atau netral [1]. Penelitian tentang analisis sentimen dapat memberikan penilaian atau evaluasi terhadap suatu produk, pelayanan dan keadaan yang sedang terjadi saat ini dari sebuah opini yang berbentuk *text*, *sms*, atau *tweet*.

Data Twitter berupa *tweet* diambil dan dilakukan *text mining*. Dengan melakukan *text mining*, dapat terlihat apakah *tweet* tersebut bernilai positif, negative atau netral. Dari hasil analisis dengan menggunakan data *tweet* dapat menghasilkan klasifikasi *tweet* positif, negative atau netral. Untuk menganalisis data *tweet* masyarakat terhadap kebijakan pemerintah Republik Indonesia terkait pencegahan penyebaran virus corona COVID-19, pada penelitian ini Bahasa Pemrograman Python digunakan untuk menganalisis data *tweet* tersebut.

Metode yang penulis gunakan untuk mengukur persentase sentimen masyarakat terhadap kebijakan pemerintah Republik Indonesia terkait upaya pencegahan penyebaran virus corona adalah *Sentiment Analysis* dengan pendekatan *Lexicon-based*.

Pendekatan berbasis leksikon (*Lexicon-based*) bergantung pada leksikon sentimen, yaitu kumpulan istilah sentimen yang diketahui dan dikompilasi sebelumnya. Sumber daya yang dapat digunakan pada analisis sentimen berbasis leksikon secara umum ada 2 jenis, yaitu berbasis kamus dan berbasis *corpus* yang menggunakan metode statistik atau semantik untuk menemukan polaritas sentiment [2]. Penulis menggunakan sumber daya kamus pada penelitian ini, dengan cara memanfaatkan library *TextBlob* pada Python yang menyediakan kamus berisi leksikon sentimen. Penelitian Analisis Sentimen yang pernah dilakukan sebelumnya dengan menggunakan data twitter dan metode

*Lexicon-Based* diantaranya adalah Adiyasa Nurfalih, Adiwijaya dan Arie Ardiyanti Suryani [3] dengan judul Analisis Sentimen Berbahasa Indonesia dengan pendekatan *Lexicon-Based* pada Media Sosial, dengan tujuan untuk menganalisa komentar masyarakat tentang produk makanan di jejaring sosial Twitter. Penelitiannya dihasilkan tingkat akurasi sebesar 66% untuk prediksi komentar positif, negative dan netral. Penelitian lainnya adalah penelitian yang dilakukan oleh Ikhwan Rustanto, and Nur aini Rakhmawati [4], dengan judul Media Sentiment Analysis of East Java Province:

*Lexicon-Based vs Machine Learning SVM*. Tujuan penelitiannya adalah menganalisa persepsi masyarakat terhadap Pemerintah Provinsi Jawa Timur (Pemrov Jawa Timur) melalui sosial media Twitter.

Metode yang digunakan *Lexicon-Based*, memperoleh akurasi sebesar 57%. Penelitian lainnya yang menggunakan metode *Lexicon Based* adalah Peneliti Rifiana Arief dan Karen Imanuel [5]. Tujuan dari penelitiannya adalah untuk mengumpulkan komentar masyarakat di sosial media Twitter tentang topik viral desa penari ke kategori positif, negative dan netral. Hasil penelitiannya diperoleh 33 tweet bernilai positif, 767 tweet bernilai netral dan 200 tweet negatif. Prosentase tweet berisi komentar positif sebesar 3.3 %, netral 76.7% dan negatif sebesar 20%.

## METODE PENELITIAN

Penulis melakukan analisa sentimen menggunakan dataset berupa tweet yang didapat dari media sosial media Twitter. Tweet diambil dari tanggal 17 April 2020 hingga 24 April 2020 berdasarkan hashtag #jokowi. Data tweet diambil menggunakan TAGS (*Twitter Archiving Google Sheet*) melalui akses Twitter API dengan cara *crawling*. Setelah mendapatkan data tweet selanjutnya dilakukan tahap *preprocessing* seperti menyeragamkan bentuk huruf pada teks menjadi huruf kecil, penghapusan URL, penghapusan mention, retweet, penggantian nama pengguna dengan spasi, *replace slang*, penghapusan tagar dan lain-lain. Setelah proses tersebut, proses selanjutnya adalah mengekstrak nilai sentimen menggunakan metode *Lexicon Based* untuk setiap katanya. Pada langkah ini penulis memanfaatkan *TextBlob library* untuk

penentuan batas ambang untuk label positif, negatif, dan netral. Tahapan berikutnya adalah penyajian hasil yang divisualisasikan dalam bentuk diagram. Tahap akhir penulis melakukan pengujian akurasi tweet dengan cara membandingkan data prediksi dengan data sebenarnya. Tahapan dalam melakukan penelitian ini dapat dilihat pada Gambar 1.

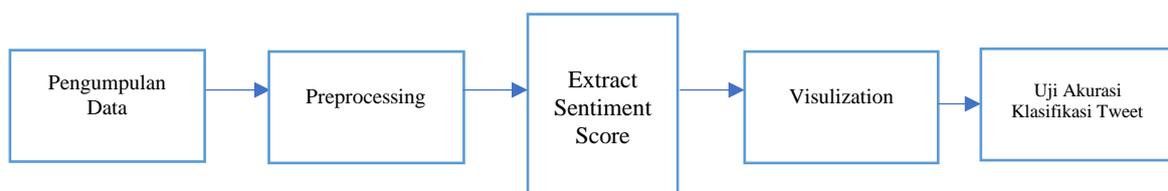
### Pengumpulan Data

Penulis melakukan pengumpulan data dimulai dari tanggal 17 April 2020 hingga 24 April 2020. Data diambil dari *micro-blogging* Twitter melalui *crawling*. *Crawling* menggunakan metode *streaming* data. Untuk memperoleh data tersebut penulis mengambil data melalui TAGS (*Twitter Archiving Google Sheet*) yang diakses melalui <https://tags.hawksey.info/get-tags/>, web yang menyediakan pengambilan data Twitter berdasarkan *hashtag*. Sebelum melakukan proses *crawling*, tahap pertama yang dilakukan adalah mendaftarkan akun twitter sebagai pengembang di situs <https://developer.twitter.com/en/apps> dan melakukan pembuatan *Apps*.

Setelah melakukan pembuatan *Apps*, didapatkan *consumer API key*, *consumer API secret key*, *access token* dan *access token secret*, hal tersebut menjadi syarat yang harus didapatkan agar dapat melakukan request pada Twitter API. Penulis mendapatkan data sebanyak 4130 tweets yang tersimpan dalam file *csv*. Dalam *crawling* tersebut menghasilkan 18 kolom yaitu *id\_str*, *from\_user*, *text*, *created\_at* *time*, *geo\_coordinates*, *user\_lang*, *in\_reply\_to\_user\_id\_str*, *in\_reply\_to\_screen\_name*, *from\_user\_id\_str*, *in\_reply\_to\_status\_id\_str*, *source*, *profile\_image\_url*, *user\_followers\_count*, *user\_friends\_count*, *user\_location*, *status\_url*, *entities\_str*.

### Preprocessing

*Preprocessing* adalah salah satu bagian terpenting dari proses analisis. Pada langkah ini, penulis menghapus atau mengganti data noise di setiap tweet, yang tidak banyak berkontribusi untuk klasifikasi sentiment [6].



**Gambar 1. Tahapan Penelitian**

Penulis melakukan 4 tahap untuk memproses tweet, yaitu *case folding*, *cleaning*, *tokenizing*, *stemming* dan *stopword removal*. Tahapan preprocessing yang dilakukan adalah sebagai berikut :

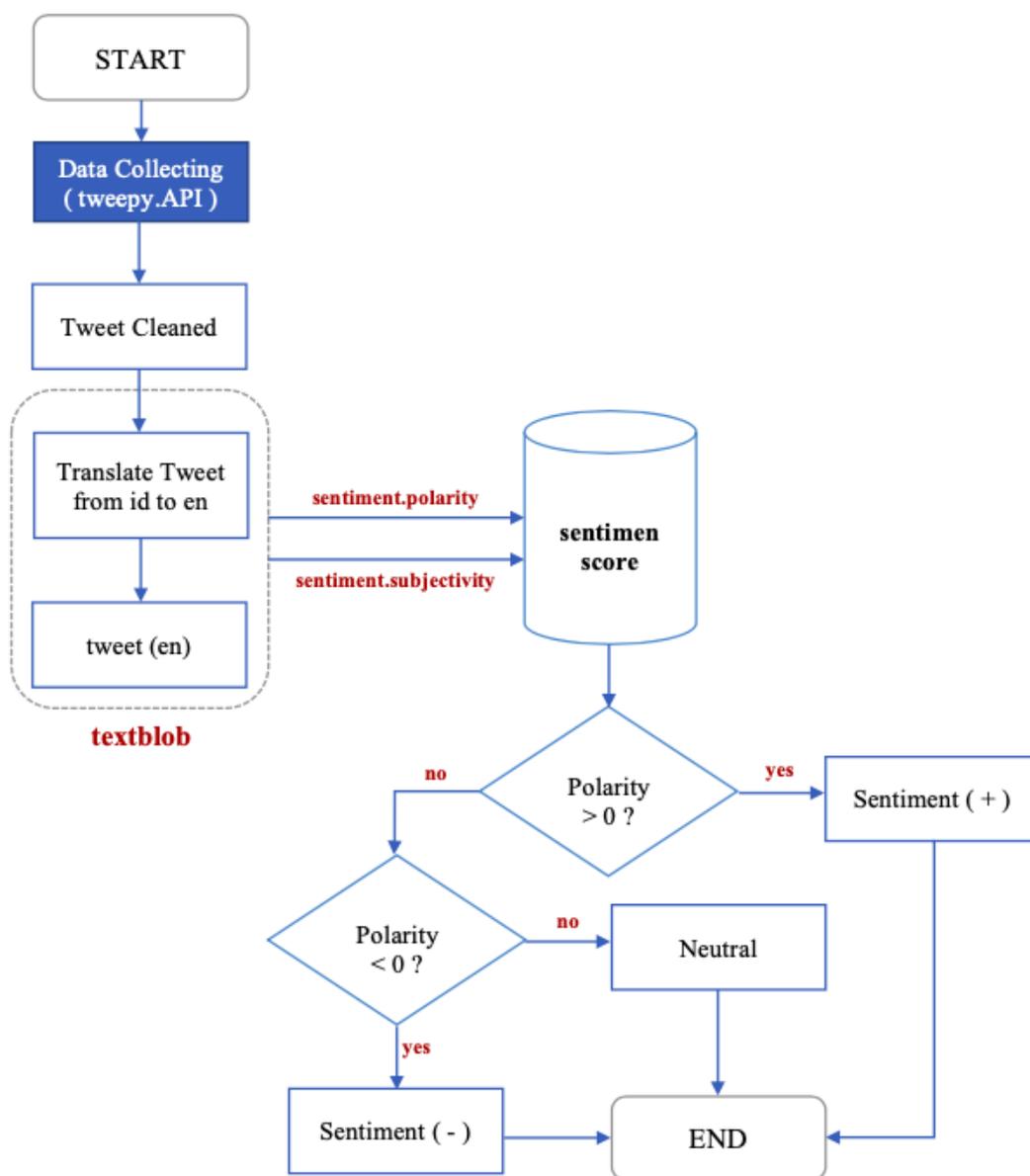
1. *Case folding* untuk menyeragamkan bentuk huruf pada teks menjadi huruf kecil (*lower case*).
2. Menghapus *mention*, *retweet*, *RT*.
3. Menghapus *hash tag*, *hyperlink*.
4. Menghapus *string* URL.
5. Menghapus karakter newline ('*\n*') yang terdapat pada teks *tweet*.
6. Menghapus tanda baca.
7. Memisahkan kata-kata menjadi kata tunggal dari sebuah kalimat.
8. Mengubah kata berimbuhan Bahasa Indonesia menjadi bentuk dasarnya menggunakan library python Sastrawi.

### **Klasifikasi Data dengan *Lexicon-Based***

Metode untuk menentukan apakah sebuah kalimat, teks atau komentar mengungkapkan sentiment positif, negatif atau netral salah satunya adalah dengan pendekatan berbasis leksikon [7]. Klasifikasi sentimen dengan Lexical Based merupakan klasifikasi yang menggunakan kamus *lexicon* Bahasa Indonesia. Namun, karena dalam penelitian ini menggunakan *python library TextBlob*, maka kamus *lexicon* yang digunakan berbahasa Inggris. Oleh karena itu teks diterjemahkan ke dalam Bahasa Inggris. *TextBlob* merupakan

*library* python yang dibangun diatas NLTK [8]. *TextBlob* berisi leksikon sentimen sebanyak 2.919 kata dengan masing-masing skor polaritas dan subjektivitas [9]. Klasifikasi ini telah dicocokkan dengan kata-kata yang terdapat dalam kamus *Lexicon* Bahasa Inggris. Jika tweets memiliki kata positif, maka akan digolongkan pada sentimen tweet positif. Jika tweets memiliki kata negatif, maka akan digolongkan pada sentimen tweet negatif. Namun pada kasus lain jika kedua kata ini bernilai sama, maka digolongkan dalam tweet netral. Diagram alur dari klasifikasi ini dapat dilihat pada flowchart di Gambar 2. Flowchart menyajikan alur dari mulai pengambilan data melalui Twitter API menggunakan TAGS. Selanjutnya data tweet akan melalui tahap preprocessing (*tweet cleaned*) lalu data tweet akan di terjemahkan dari Bahasa Indonesia ke dalam Bahasa Inggris dengan menggunakan *python library TextBlob*.

*Python library TextBlob* selain dapat digunakan untuk menerjemahkan kata, dapat juga digunakan untuk proses analisa sentiment. Hasil terjemahan tweet Berbahasa Inggris selanjutnya dilakukan proses analisa sentimen. Klasifikasi menggunakan metode *Lexicon Based* menggunakan *python library TextBlob* dari data tweets sebanyak 4130 menghasilkan sentimen positif sebanyak 1942, sentimen negatif sebanyak 1259 dan sentimen netral sebanyak 929.



**Gambar 2. Flowchart Analisis Sentimen**

**Extract Sentiment Score**

Hasil *preprocessing* pada tahap sebelumnya menghasilkan berupa kumpulan kata sifat, kata keterangan, kata benda, dan kata kerja dalam Bahasa Inggris.

Penulis mengekstrak nilai sentimen menggunakan metode *Lexicon Based* untuk

setiap katanya. Pada langkah ini penulis memanfaatkan TextBlob library pada python. Untuk penentuan batas ambang untuk label positif, negatif, dan netral ditunjukkan pada Gambar 3. Batas ambang untuk label positif adalah jika  $score > 0$ , label negatif jika  $score < 0$  dan netral jika  $score = 0$ .

```

def sentimentlabels(senscores):
labels = []
for score in senscores:
    if(score > 0):
        labels.append('positive')
    if(score < 0):
        labels.append('negative')
    else:
        labels.append('netral') return labels

```

**Gambar 3. Algoritma untuk Mengekstrak Nilai Sentimen**

### Uji Akurasi Klasifikasi Tweet

Uji klasifikasi sentimen dilakukan dengan membandingkan data prediksi dengan data sebenarnya.

Berbeda dengan data prediksi dalam format hasil klasifikasi yang dihasilkan oleh algoritma yang digunakan, data aktual dihasilkan dalam format hasil klasifikasi yang dihasilkan oleh manusia menggunakan pelabelan manual [10]. Textblob bersifat unsupervised, maka tidak dibutuhkan data latih untuk mengklasifikasi sentimen. Pengujian dalam penelitian menggunakan pengujian akurasi. Akurasi mewakili perbandingan jumlah kasus yang diklasifikasikan dengan benar dengan jumlah semua kasus. Untuk menghitung akurasi klasifikasi digunakan metode confusion matrix atau dalam unsupervised learning biasa disebut dengan matching matrix.

*Confusion Matrix* adalah *tools* evaluasi visual yang digunakan dalam *machine learning*. Kolom *Confusion Matrix* mewakili hasil kelas prediksi, dan baris mewakili hasil kelas yang sebenarnya [11]. Tabel matriks

terdiri dari data aktual dan prediksi, juga dikenal sebagai model confusion matrix 3x3, atau confusion matrix yang diperluas, seperti yang ditunjukkan pada Tabel 1. Perhitungan akurasi ditunjukkan pada persamaan 1.

*Akurasi*

$$= \frac{AA + BB + CC}{AA + AB + AC + BA + BB + BC + CA + CB + CC} \quad (1)$$

## HASIL DAN PEMBAHASAN

### Hasil Pengumpulan Data

Berdasarkan uraian sebelumnya, data tweet didapat dengan menggunakan teknik *crawling*. Penulis mendapatkan data sebanyak 4130 tweets yang tersimpan dalam file *csv*. Proses *Crawling* dilakukan menggunakan TAGS. Gambar 4 menunjukkan tampilan dari web TAGS. Hasil *crawling* ditunjukkan pada Gambar 5. Tabel 1. menunjukkan dataset yang digunakan. TAGS digunakan untuk proses pengumpulan data yang diambil melalui website <https://tags.hawksey.info/get-tags/> dengan menggunakan Twitter API. Data secara otomatis akan dikumpulkan berdasarkan *hashtag* dan tanggal pengambilan

data. Seperti yang sudah dijelaskan sebelumnya, *crawling* data tweet dilakukan menggunakan *hashtag* Jokowi dengan periode tanggal 17 April 2020 hingga 24 April 2020.

*Crawling* tersebut menghasilkan 18 kolom yaitu *id\_str*, *from\_user*, *text*, *created\_at*, *time*, *geo\_coordinates*, *user\_lang*, *in\_reply\_to\_user\_id\_str*,

*in\_reply\_to\_screen\_name*, *from\_user\_id\_str*, *in\_reply\_to\_status\_id\_str*, *source*, *profile\_image\_url*, *user\_followers\_count*, *user\_friends\_count*, *user\_location*, *status\_url*, *entities\_str*. Gambar 5 menunjukkan 5 kolom pertama hasil *crawling* data tweet menggunakan TAGS.

**Tabel 1. Tabel Confusion Matrix**

		Data Prediksi		
		Kelas A	Kelas B	Kelas C
Data Aktual	Kelas A	AA	AB	AC
	Kelas B	BA	BB	BC
	Kelas C	CA	CB	CC



**Gambar 4. TAGS-Twitter Archiving Google Sheet**

	A	B	C	D	E
1	id_str	from_user	text	created_at	time
2	1253835459471372	LohmenzGonjreng	Oh mengapa? Ada apa sebenarnya? #Jokowi2Periode #JokowiMaruf2019 #Jokowi #Po	Fri Apr 24 23:57:23	25/04/2020 00:57:23
3	1253832072461824	AsmaraRomy	Gimana kabarnya. Programnya #gajelas #kartuprakerja #Jokowi .	Fri Apr 24 23:43:55	25/04/2020 00:43:55
4	1253786938626228	eka_adhie	sangat disayangkan tim bapak jokowi sgt lamban dikala covid. tim lambat atau sengaja	Fri Apr 24 20:44:35	24/04/2020 21:44:35
5	1253744607449329	eka_adhie	Nah bener tu pak wi.. Tp kok ga dr kmaren? Tim anda perlu digeser deh.. Masak kalah	Fri Apr 24 17:56:22	24/04/2020 18:56:22
6	1253731427197149	Lohmenz	Ya Tuhan... apa yang telah terjadi di negeri ini? #Jokowi #JokowiLempatHandukSaja #jok	Fri Apr 24 17:04:00	24/04/2020 18:04:00
7	1253724872359854	monitorindo	Ditelepon Jokowi, Presiden Trump Segera Kirim Ventilator Corona ke Indonesia: Mafia	Fri Apr 24 16:37:57	24/04/2020 17:37:57
8	1253722250009665	rayreign	RT @detikcom: "Kalau itu bukan mudik, itu namanya pulang kampung. Memang bekerj	Fri Apr 24 16:27:32	24/04/2020 17:27:32
9	1253711075670781	RizaMillati	#Jokowi duh selama ini aku kurang paham dengan bahasa indonesia https://t.co/WYrB	Fri Apr 24 15:43:07	24/04/2020 16:43:07
10	1253710645721133	RizaMillati	RT @detikcom: "Kalau itu bukan mudik, itu namanya pulang kampung. Memang bekerj	Fri Apr 24 15:41:25	24/04/2020 16:41:25
11	1253701983963238	newsmerahputih	COVID-19 Bikin Puasa Tahun Ini Beda, Jokowi: Tidak Ada Kesemaranan di Jalanan. #n	Fri Apr 24 15:07:00	24/04/2020 16:07:00

**Gambar 5. Hasil Crawling Menggunakan TAGS**

**Tabel 2. Contoh Dataset yang Digunakan**

<b>id_str</b>	<b>from_user</b>	<b>text</b>	<b>create_at</b>
1250541782279966720	edho_doink	PSBB !! apakah penerapatn kebijakan ini optimal dan dipatuhi ? pantauan lapangan menjawab TIDAK , karna apa? karena kebutuhan pangan masyarakat terutama miskin tidak terpenuhi oleh sebab itu aktifitas akan trs berlanjut demi perut keluarga .#Jokowi	Wed Apr 15 21:49:29 +0000 2020
1251400057682751490	Lewisamuelartha	Mantap bapak Presiden Jokowi Saya sangat kagum dengan apa yang bapak kerjakan saat ini #JokowiMaruf2019 #Jokowi #banyakkerja #KerjaKerjaKerja #majuIndonesia <a href="https://t.co/E21CBRfi6Y">https://t.co/E21CBRfi6Y</a>	Sat Apr 18 06:39:58 +0000 2020
1253178435398074370	nursslmh_	Kebayang gasi ngindarin biar ga meninggal gara gara covid19 sampe diberlakukannya PSBB, tapi meninggal gara gara kelaparan. #COVID19 #COVID19indonesia #Jokowi #prabowo	Thu Apr 23 04:26:36 +0000 2020

Kolom *from\_user* menunjukkan nama user, kolom *text* berisikan data tweet dan kolom *created\_at* dan *time* menunjukkan waktu tweet tersebut di *posting* ke media sosial Twitter.

Hasil *crawling* didapatkan 4130 tweets dan tersimpan dalam file *csv*. Selanjutnya hasil pengumpulan data tweet tersebut dijadikan dataset dan digunakan dalam proses analisa sentimen.

Tabel 2 menunjukkan contoh dataset

yang penulis gunakan untuk melakukan proses analisa sentimen. Contoh dataset yang disajikan dalam Tabel 1 yang digunakan untuk proses preprocessing adalah data pada kolom *text*.

### **Hasil Preprocessing**

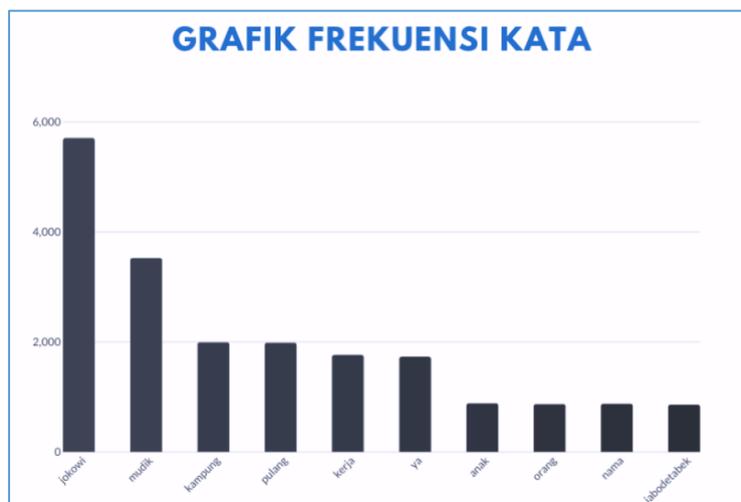
Setelah tahap pengumpulan data, tahap selanjutnya adalah preprocessing. Berdasarkan uraian sebelumnya tahap ini terdiri dari *case folding*, *cleaning*, *tokenizing*, *stemming* dan

*stopword removal* [12]. Hasil tahap preprocessing secara garis besar ditunjukkan pada Tabel 3 dan Gambar 7. Tabel 3 menyajikan hasil dari tahap preprocessing data dan Gambar 7 menunjukkan persebaran frekuensi 10 kata terbanyak setelah dilakukan

tahap preprocessing. Contoh data tweet yang sudah melalui proses preprocessing tersaji dalam kolom 2 di Tabel 3. Data tweet sudah tidak mengandung RT, menghilangkan username, menghilangkan tag, menghilangkan link URL dan lain-lain.

**Tabel 3. Contoh Hasil Data Processing**

Data Sebelum Preprocessing	Data Sesudah Preprocessing
RT @Lewisamuelartha: Mantap bapak Presiden Jokowi Saya sangat kagum dengan apa yang bapak kerjakan saat ini #JokowiMaruf2019 #Jokowi #banyakkerja #KerjaKerjaKerja #majuIndonesia <a href="https://t.co/E21CBRfi6Y">https://t.co/E21CBRfi6Y</a>	mantap bapak presiden jokowi saya sangat kagum dengan apa yang bapak kerja saat ini jokowi maruf 2019 jokowi banyak kerja kerja kerja kerja maju indonesia
Ditunda untuk sementara Hingga semua seperti sedia kala	ditunda untuk sementara hingga semua seperti sedia kala
#dirumahaja #jokowi <a href="https://t.co/EINnjclwpT">https://t.co/EINnjclwpT</a>	
LANGKAH JOKOWI HADAPI WABAH VIRUS CORONA DIPUJI, EPIDEMIOLOGIS INGGRIS: TAK BISA SENANGKAN SEMUA ORANG	langkah jokowi hadapi wabah virus corona dipuji epidemilogis inggris tak bisa senang semua orang
#Jokowi #VirusCorona #Covid19 #Jokowidodo <a href="https://t.co/ImCtk7XWSJ">https://t.co/ImCtk7XWSJ</a> via @tribunkaltim	



**Gambar 7. Grafik Frekuensi Kata**

**Tabel 4. Contoh Hasil *Sentiment Score***

Hasil preprocessing	Hasil translate	Sentimen Score	Sentimen Label
mantap bapak presiden jokowi saya sangat kagum dengan apa yang bapak kerjakan saat ini jokowi maruf 2019 jokowi banyak kerja kerja kerja kerja maju indonesia	great mr. president jokowi, i am very impressed with what you are currently doing. jokowi maruf 2019 jokowi work work work work advanced indonesia	0,55	positive
ditunda untuk sementara hingga semua seperti sedia kala	temporarily postponed until everything is as it was before	0,0	neutral
corona lockdown narapidana bebas negaraku kacau dan pemerintah seperti kata semua bahwa buat marah masyarakat adalah hal untung	corona lockdown free convicts, my country is chaotic and its government seems to say everything that makes people angry is beneficial	-0,5	negative
langkah jokowi hadapi wabah virus corona dipuji epidemiologis inggris tak bisa senangkan semua orang	Jokowi's step in dealing with the corona virus outbreak was praised, the English epidemic could not please everyone	0,0	neutral

Sebelum masuk ke tahap analisa sentimen, data hasil dari tahap preprocessing dapat menghasilkan kata yang sering muncul. Data tersebut dapat dijadikan indikator bagi penulis sebagai acuan relevansi antara kata yang sering muncul dengan hashtag yang digunakan pada saat pengumpulan data, yaitu #jokowi. Grafik pada Gambar 7 menunjukkan kata Jokowi sebagai kata yang sering muncul, dengan kisaran angka 5800.

### Hasil Analisa Sentimen

Pada tahapan *Extract Sentimen Score* dihasilkan label sentimen dalam setiap kata teks. Hasil ditunjukkan pada Tabel 4 berupa contoh hasil preprocessing, sentimen score dan sentimen label.

Tabel 4 menunjukan hasil translate data preprocessing yang menggunakan Bahasa Indonesia ke dalam Bahasa Inggris, *sentiment score* dan *sentiment label* dengan menggunakan *python library textblob*. Batas ambang

untuk label positif adalah jika score > 0, label negatif jika score < 0 dan netral jika score = 0. Contoh score Tabel 4 pada baris 1 adalah 0,55, score 0,55 termasuk dalam label positif karena batas ambang nilai score > 0. Contoh score Tabel 4 pada baris 2 adalah 0,0, score 0,0 termasuk dalam label netral karena batas ambang nilai score = 0. Contoh score Tabel 4 pada baris 3 adalah -0,55, score -0,55 termasuk dalam label negatif karena batas ambang nilai score < 0.

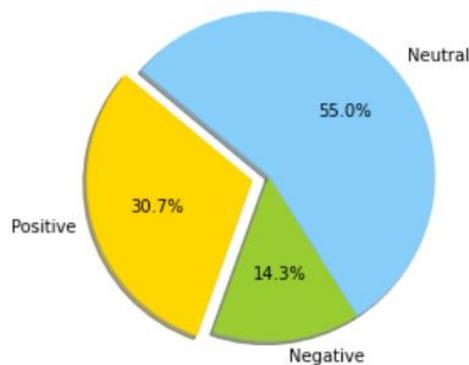
### Hasil Visualisasi

Berdasarkan data dan analisis sentimen *lexicon-based*, maka dihasilkan visualisasi

dalam bentuk *pie chart*, seperti yang ditunjukkan pada Gambar 8 bahwa terdapat 30,7% tanggapan positif terhadap kebijakan perintah dalam upaya pencegahan penyebaran virus corona COVID-14, 3% tanggapan negatif, dan 55,0% netral.

### Pengujian Klasifikasi

Hasil klasifikasi menggunakan TextBlob menghasilkan kelas sentimen positif lebih banyak dari hasil pelabelan manual, jumlah kelas sentiment positif dari klasifikasi TextBlob berjumlah 370 sementara pelabelan manual menghasilkan klasifikasi 347 teks.



Gambar 8. Hasil Visualisasi Analisa Sentimen

Tabel 5. Tabel *Confusion Matrix*

		Data Prediksi		
		Positif	Netral	Negatif
Data Aktual	Positif	195	38	114
	Netral	20	114	17
	Negatif	155	21	532

$$Akurasi = \frac{195 + 114 + 532}{1026} = 0.819 = 81.9\%$$

Di kelas sentimen negatif, TextBlob mengklasifikasikan 173 teks, dibandingkan pelabelan manual yang dapat mengklasifikasikan 151 teks sebagai sentiment negatif. Berbeda dengan kelas netral, TextBlob menghasilkan klasifikasi netral sebanyak 663, lebih sedikit dibandingkan pelabelan manual yang mengklasifikasikan 708 teks sebagai teks netral.

Berdasarkan perhitungan pada Tabel 5, diketahui hasil klasifikasi sentiment menggunakan textblob memberikan akurasi secara keseluruhan sebesar 81,9 %. Nilai akurasi mengindikasikan seberapa akurat sistem dapat mengklasifikasikan data dengan benar.

## KESIMPULAN DAN SARAN

Penelitian ini telah menghasilkan analisis sentimen berdasarkan tanggapan masyarakat terhadap kebijakan pemerintah dalam upaya pencegahan penyebaran virus corona COVID-19. Metode *Lexicon Based* dapat diterapkan dalam teks berbahasa Indonesia dengan terlebih dahulu dilakukan proses penerjemahan ke dalam Bahasa Inggris. Berdasarkan persentase tanggapan masyarakat terhadap kebijakan pemerintah dalam upaya pencegahan penyebaran virus corona COVID-19 ini dapat disimpulkan bahwa sebagian besar masyarakat Indonesia memberikan tanggapan positif.

Berdasarkan hasil pengujian yang dilakukan didapatkan tingkat akurasi 81.9 %,

karena dalam menganalisis sentiment teks antara algoritma dan manusia berbeda. Manusia memiliki emosi dalam memberikan ukuran sentiment dalam sebuah teks. Saran penelitian selanjutnya adalah dilakukannya perbandingan *score* dengan menggunakan *VADER* sentiment.

## DAFTAR PUSTAKA

- [1] L. N. Pradany and C. Faticah, "Analisa Sentimen Kebijakan Pemerintah Pada Konten Twitter Berbahasa Indonesia Menggunakan SVM dan K-Medoid Clustering," *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, vol. X1, no. 1, pp. 59-66, 2016.
- [2] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications:," *Ain Shams Engineering Journal*, vol. 1, pp. 1093-1113, 2013.
- [3] N. Adiyasa , Adiwijaya and A. . S. Ardiyanti, "Analisis Sentimen Berbahasa Indonesia dengan pendekatan Lexicon-Based pada Media Sosial," vol. 2, no. 1, January - March 2017.
- [4] I. Rustanto and N. a. Rakhmawati, "Media Sentiment Analysis of East Java Province: Lexicon-Based vs Machine Learning SVM," *IPTEK Journal of Proceedings Series*, no. 6, pp. 203-208, 2020.
- [5] A. Rifiana and K. Imanuel, "Analisis Sentimen Topik Viral Desa Penari Pada

- Media Sosial Twitter Dengan Metode Lexicon Based," *Jurnal Ilmiah MATRIK*, vol. 21, no. 3, 2019.
- [6] H. M. Keerthi Kumar and B. S. Harish, "Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation," vol. 3, pp. 19-30, 2017.
- [7] K. . M. Nahar, A. Jaradat, M. S. Atoum and F. Ibrahim, "Sentiment Analysis And Classification Of Arab Jordanian Facebook Comments For Jordanian Telecom Companies Using Lexicon-Based Approach And Machine Learning," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 06, no. 03, pp. 247-262, 2020.
- [8] W. N. S. Wan Min and N. Z. Zulkarnain, "Comparative Evaluation of Lexicons in Performing Sentiment Analysis," *Journal Of Advanced Computing Technology And Application (JACTA)*, vol. 2, no. 1, pp. 14-20, 2020.
- [9] S. Loria, *textblob Documentation Release 0.16.0*, 2020.
- [10] U. Khaira, R. Johanda, U. P. Pradita Eko and T. Suratno, "Sentiment Analysis of Cyberbullying on Twitter," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)*, vol. 3, no. 1, pp. 21-26, 2020.
- [11] J. Xu, Y. Zhang and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Information Sciences*, vol. 507, no. 0020-0255, pp. 772-794, 2020.
- [12] G. A. Buntoro, T. B. Adji and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *CITEE Universitas Gajah Mada*, pp. 39-43, 2014.