

## Analisis Sentimen Pandemi Covid-19 Pada *Streaming* Twitter Dengan *Text Mining* Python

Anni Karimatul Fauziyyah<sup>1)</sup>, Deden Hardan Gautama<sup>2)</sup>

<sup>1, 2)</sup> Program Studi Informatika, Universitas Alma Ata

<sup>1)</sup> anni.karim@almaata.ac.id, <sup>2)</sup> hardan@almaata.ac.id

### ABSTRACT

*The impact of the novel coronavirus (COVID-19) is widespread and will likely shape community behavior for months to come. And while the humanitarian and safety-related aspects of this outbreak are top of mind globally, it's unquestionable that social distancing, quarantining, and staying home will have a significant effect on media consumption, which could rise up to 60%, according to recent research from Nielsen's U.S. media team. Social media, now a part of everyday life for most consumers engaged with the world digitally, became the primary source for buzz about all things COVID-19 as worries and news intensified. Sentiment analysis is applied in this study to analyze the opinions, feelings, and interests of individuals in the COVID-19. The purpose of this study is to analyze sentiment based on an opinion by classifying individual feelings such as sadness, happiness, or panic in facing a COVID-19 into sentiment level that is negative, positive or, neutral. In this paper, an open-source approach is presented where we have collected tweets from the Twitter API and then reprocessing, analyzing and, visualizing these tweets using python. Furthermore, Twitter data streaming will be processed and cleaned to parse data that can be classified based on opinion with a text mining algorithm using text blob Python. Feature extraction is done for the relationship between words by the Bigram and N-gram methods.*

**Keywords:** *Sentiment analysis, COVID-19, text mining, Bigram, Ngram*

### I. PENDAHULUAN

Di era semakin berkembangnya teknologi internet pada saat ini pertumbuhan jumlah data digital semakin besar. Media sosial adalah salah satu contoh penghasil data internet terbesar (Dhawan, 2014). Lebih dari jutaan orang atau organisasi menggunakan media sosial sebagai bentuk eksistensinya di dunia maya (Sivarajah, 2019). Menjalinkan hubungan sosial antar pengguna, berbagi informasi dan peristiwa akan menghasilkan data besar yang *real time*.

Seperti peristiwa pandemi COVID-19 atau virus corona yang merebak di seluruh dunia pada awal tahun 2020. Berdasarkan (JHU, 2020) pesebaran kasus COVID-19 di seluruh dunia pada 15 Maret 2020 sudah mencapai 1,986,986 kasus terkonfirmasi dan kasus meninggal mencapai 126,812. Sedangkan di Indonesia tercatat telah mencapai 4,839 kasus per 15 April 2020 dengan kasus meninggal 459 (BNPB, 2020). Dampak dari COVID-19 tersebar luas dan memungkinkan membentuk perilaku masyarakat selama masa pandemic belum berakhir. Sementara aspek-aspek yang terkait dengan kemanusiaan dan keselamatan dari pandemi ini menjadi perhatian utama secara global, dimana pembatasan jarak sosial, karantina, dan anjuran tinggal di rumah akan memiliki efek signifikan pada konsumsi media, yang bisa naik hingga 60%, menurut penelitian terbaru dari tim media AS (Nielsen's U.S, 2020). Media sosial, sekarang menjadi bagian dari kehidupan sehari-hari bagi sebagian besar masyarakat yang terlibat dengan dunia secara digital, dan menjadi sumber utama untuk semua hal tentang COVID-19 ketika kekhawatiran dan berita semakin intensif. Akibat dari pandemi ini sejumlah negara di seluruh dunia memberlakukan *lockdown* dan pembatasan jarak social yang berimbas pada aktivitas masyarakat sehingga lebih aktif bersosialisasi di dunia maya dengan internet.

Media sosial terutama twitter semakin menggaungkan peristiwa pandemi COVID-19 di jagad maya dengan munculnya *hashtag* tentang peristiwa pandemi COVID-19 hingga menjadi *trending* topik. Data yang dihasilkan twitter akan menjadi penting dan bermanfaat bagi masyarakat atau organisasi jika diolah dan di analisa. Salah satunya analisis sentimen twitter, merupakan metode untuk mengeksplorasi apakah konten yang dikumpulkan dalam keadaan positif, negatif atau netral (Luqyana, 2018). Analisis sentimen diterapkan pada penelitian ini untuk menganalisa opini, perasaan, pandangan, dan minat individu pada peristiwa pandemi COVID-19 dengan *streaming* data pengguna twitter yang membahas tentang topik COVID-19. Merebaknya pandemi ini diseluruh dunia dan berlakunya pembatasan interaksi sosial berpengaruh pada kondisi sosial masyarakat beredarnya isu-isu baik positif maupun negative bahkan timbul kepanikan besar di media social (Martin, 2020).

Berdasarkan uraian diatas, maka penulis akan menerapkan analisis sentimen berdasarkan dari opini dengan mengklasifikasikan perasaan individu seperti sedih, senang, atau panik menghadapi pandemic COVID-19 kedalam level sentimen yaitu negatif, positif atau netral. Tidak semua opini dapat digunakan sebagai analisa, sehingga harus dilakukan pemilahan, seperti pada penelitian (Setiowati, 2016) hanya opini positif dan negatif yang digunakan untuk menganalisa pengaruh opini di sosial media pada tingkat pelayanan hotel dan travel. Saat ini bahasa pemrograman untuk analisa data mining sangat berkembang. Pada penelitian (Saini, 2019) menggunakan pemrograman R untuk menganalisa sentimen data twitter. Sedangkan pada penelitian ini menggunakan python dengan library NLTK untuk *artificial intelligence*, yang lebih produktif daripada bahasa konvensional untuk pemecahan masalah yang melibatkan manipulasi *string* dan pencarian di library (Adawadkar, 2017).

Tujuan dari penelitian ini adalah melakukan *collecting* data twitter secara *real time* atau *streaming* yang diperlukan untuk analisa sentimen sesuai *trend* pandemi COVID-19. TwitterAPI digunakan untuk *streaming* data *tweet* dari Twitter dan menyimpan jumlah *tweet* secara *real time*. Selanjutnya data *streaming* twitter akan di *processing* dan *cleaning* untuk memilah data yang dapat diklasifikasikan berdasarkan opini dengan algoritma data *mining* dengan pemrograman python sehingga dapat dilakukan analisa dan visualisasi *tweet* ini menggunakan python. Ekstraksi fitur dilakukan untuk hubungan antara kata-kata dengan metode Bigram dan N-gram digunakan untuk analisis sentimen berdasarkan dari opini dengan mengklasifikasikan perasaan individu seperti sedih, senang, atau panik menghadapi pandemic COVID-19 kedalam level sentimen yaitu negatif, positif atau netral.

## II. TINJAUAN PUSTAKA

### 2.1 Text Mining

*Text mining* dan *data mining* di persepsikan dalam keilmuan yang sama di karenakan algoritma sama dapat digunakan dalam kedua konsep *mining* tersebut (Eroglu, 2018). Namun, keduanya berbeda dalam arti bahwa *data mining* melibatkan data terstruktur, sementara teks berkaitan dengan fitur-fitur tertentu dan relatif tidak terstruktur dan memerlukan *preprocessing*. *Text mining* adalah bidang yang saling terkait dengan *Natural Language Processing* (NLP). Berikut proses utama *Text mining* berdasarkan (Yogapreethi, 2016):

- Ekstrasi Informasi  
Langkah awal untuk menganalisis teks yang tidak terstruktur dengan menggali informasi. Proses pencarian dan hubungan antar frasa, untuk ukuran teks yang besar.
- *Clustering*  
Memisahkan teks menjadi satu kelompok dan dengan cara memberi bobot untuk setiap kata. Daftar kelas dihasilkan dengan algoritma *clustering* untuk menghitung *similarity*.

- **Klasifikasi**  
Menemukan topik utama dokumen dengan menentukan meta data berupa label kelas untuk menganalisis dokumen.
- **Visualisasi informasi**  
Menampilkan hasil mengekstraksi pola informasi, digunakan untuk interaksi pengguna dengan dokumen.

## 2.2 Streaming Data Twitter

Data twitter menjadi salah satu sumber data penelitian yang lama sering digunakan, (Cheng, 2013) menjelaskan mengapa Twitter digunakan sebagai sumber informasi: data dari berbagai sumber, melampirkan konten, komunikasi *real time*. Komponen *tweet* yang dapat digunakan untuk mengekstrak informasi:

- *User name*: identifikasi pengguna
- *Time Stamp*: waktu saat *tweet* itu dikirim
- *Tweet Text*: isi dari text *tweet* - harus kurang dari 140 karakter.
- *Hashtags*: symbol #, tagar dikaitkan dengan topik tertentu.
- *Replies*: balasan pesan teks dari *tweet*
- *Retweet* : ketika pengguna membagikan *tweet* dengan pengikut mereka.

Untuk *streaming* data *tweet* dapat di akses dengan *API REST Twitter* yang memungkinkan untuk mengakses *tweet* semua orang. Tetapi API memiliki batasan hanya 6 sampai 9 hari terakhir akses *tweet* dan 18.000 *tweet* dalam satu panggilan. Pada python memiliki library untuk akses twitter salah satunya *tweepy*.

## 2.3 Data Collection dan Cleaning di Python

Data *collection* merupakan proses pengumpulan dan pengukuran informasi tentang variabel yang ditargetkan dengan cara sistematis. Pengumpulan data yang akurat sangat penting dalam menjaga integritas penelitian. Data *cleaning* proses identifikasi informasi yang tidak lengkap, atau tidak relevan dan kemudian memodifikasi atau menghapus data yang kotor (Saini, 2019).

### **TextBlob**

*Library* python untuk memproses data tekstual. Menyediakan API yang konsisten untuk menganalisa ke dalam tugas-tugas pemrosesan bahasa alami (NLP) yang umum seperti *part-of-speech tagging*, ekstraksi kata benda, analisis sentimen (Praneeth, 2020). Skema *library Textblob* : `Class textblob.blob.BaseBlob(text, tokenizer=None, pos_tagger=None, np_extractor=None, analyzer=None, parser=None, classifier=None, clean_html=False)`

Kelas dasar abstrak yang akan diwarisi oleh semua kelas *textblob*. Termasuk kata-kata, tag POS, NP, dan properti jumlah kata, termasuk metode dasar *dunder* dan *string* untuk membuat objek seperti *string* Python.

## 2.4 Analisis Sentimen

Langkah analisis sentimen dapat diterapkan sebagai berikut (Kiruthika, 2016):

- **Level 1**  
Level kalimat: Mencari sentimen netral, negatif dan positif untuk setiap baris.
- **Level 2**  
Level dokumen: menganalisa sentimen seluruh dokumen sebagai netral atau negatif atau positif .
- **Level 3**  
Level aspek: Beberapa analisis sentimen menerapkan pengelompokan pada level ini di mana semua atribut yang memiliki hasil sentimen yang sama dikumpulkan.

- Level 4

Level pengguna: memanfaatkan visualisasi data dari analisis sentiment untuk berinteraksi dengan pengguna social.

Dengan Python, metode *Textblob.ngrams* () mengembalikan daftar tupel kata-kata berurutan. Ini dapat digunakan untuk mendapatkan Unigrams (kata tunggal), Bigrams (dua kata) atau n-gram (beberapa kata berturut-turut).

Unigram: *Textblob.ngrams* (n = 1) Bigram: *Textblob.ngrams* (n = 2)

N-gram: *Textblob.ngrams* (n = 3 atau lebih).

## 2.5 Pandemi COVID-19

Pada 9 Januari 2020, CDC China melaporkan coronavirus baru sebagai agen penyebab wabah, yang masih satu jenis dengan SARS-CoV, disebut sebagai penyakit coronavirus baru 2019 (COVID-19). Pada 11 Maret 2020, 118.598 kasus COVID-19 dilaporkan di seluruh dunia oleh lebih dari 100 negara. Belum di temukan vaksin, mayoritas kasus (80%) adalah infeksi saluran pernapasan dan pneumonia (ECDC, 2020). Tindakan preventif kesehatan masyarakat untuk mengurangi dampak pandemi salah satunya menjalankan aturan *social distancing*. Kecerdasan teknologi media sosial dimanfaatkan untuk meningkatkan untuk menjalani *social distancing* dengan media sosial berkaitan dengan situasi epidemic dapat digunakan sebagai sumber informasi untuk melakukan mobilisasi masyarakat mengikuti prosedur karantina, dengan mengurangi situasi panik dan ketakutan (Martin, 2020).

## III. METODE PENELITIAN

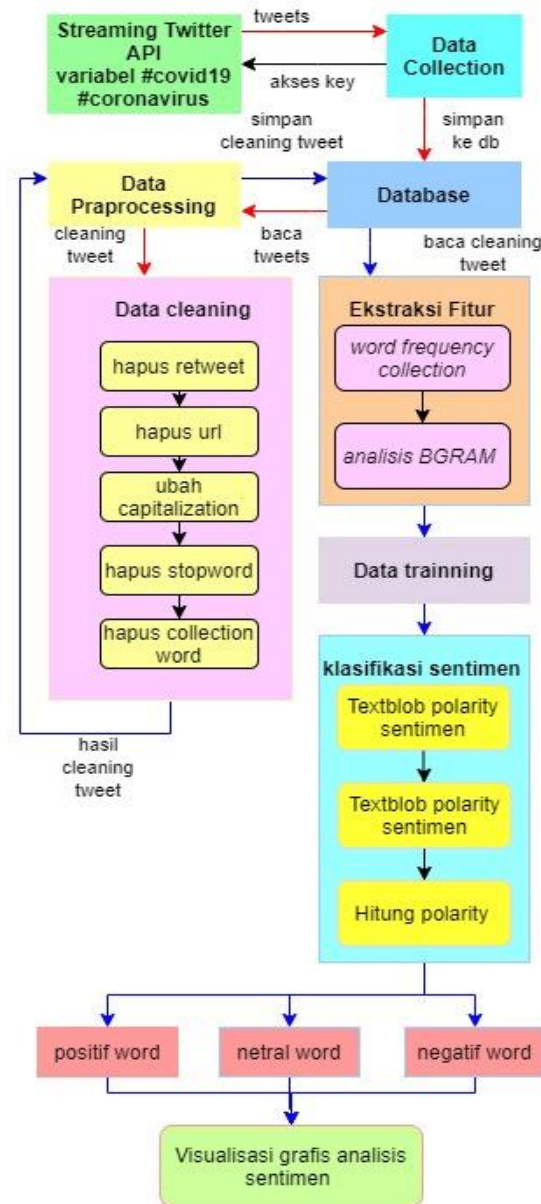
Pada penelitian ini model penelitian yang di ajukan ditunjukkan pada Gambar. 1. Pada tahapan awal adalah proses *streaming* data API twitter, untuk mengakses diperlukan 4 key yaitu: *consumer key*, *consumer secret key*, *access token key*, *access token secret key*. Kata pencarian yang digunakan adalah variable #coronavirus #COVID-19 dengan pengaturan *tweet* bahasa inggris dan tanggal dari 6 april 2020. Pada penelitian ini akan di lakukan data *collection* 4000 *tweet* dari *streaming* twitter untuk menganalisis sentimen pengguna twitter dengan variabel pandemi COVID-19, coronavirus menggunakan python. Selanjutnya data *collection* disimpan ke database nosql mongodb.

Data *collection* yang disimpan di database dilakukan proses data *cleaning* sebelum data di analisa.

- *filter:retweet s* : menghapus *retweet* yang mengandung konten duplikat,
- *re.sub* : menghapus url (tautan) menggunakan ekspresi reguler yang diakses dari kumpulan karakter yang ditentukan,
- *set(lower\_case)*:mengubah kata menjadi huruf kecil untuk menghilangkan kapitalisasi kata,
- *nlk*: menghapus stopword daftar kata yang kurang bermanfaat seperti *who*, *you*, *is*, *am* *are*

Proses ekstraksi fitur mencari kumpulan kata dari *tweet* yang telah melalui proses *cleaning* dimulai dengan menganalisa 100 kata teratas yang sering muncul pada *tweet* . *collection.Counter*:menampilkan kata-kata yang paling umum digunakan dan berapa kali digunakan. Selanjutnya dihitung nilai hubungan antar 100 kata teratas dengan bigrams: menghitung hubungan antar 2 kata.

Klasifikasi analisis sentiment menggunakan properti sentimen dari *Textblob* mengidentifikasi sikap dalam data teks tentang subjek yang menarik dalam bentuk Sentimen (polaritas, subjektivitas). Skor polaritas adalah float di kisaran -1.0 dan 1.0. Nilai lebih dekat ke 1 menunjukkan lebih banyak positif, sedangkan nilai lebih dekat ke -1 menunjukkan lebih banyak negatif.



Gambar 1. Alur sistem analisis sentimen twitter python

#### IV. HASIL DAN PEMBAHASAN

Hasil penelitian analisis sentimen terhadap *streaming* data twitter dengan python. Pada penelitian ini akan di lakukan data *collection* 4000 *tweet* dari *streaming* twitter untuk menganalisis sentimen pengguna twitter dengan variabel pandemi COVID-19, coronavirus menggunakan python. Kata pencarian yang digunakan adalah variable *#coronavirus* *#COVID-19* dengan pengaturan *tweet* bahasa inggris dan data dari tanggal 6 april 2020. Selanjutnya data *collection* disimpan ke database nosql mongodb.

Pada Gambar 2 fungsi di python untuk *collection* data *tweet* dengan *keyword* pencarian: *#COVID-19*, *#coronavirus*, *#coronavirusoutbreak* dan mengambil data *user screen name* dengan fungsi *tweet .user.screen\_name*, untuk lokasi dengan fungsi *tweet .user.location* serta *tweet text* untuk status *tweet* .

```
# Collect tweets
tweets = tw.Cursor(api.search,
                    q=new_search,
                    lang="en",
                    since=date_since).items(4000)

status_text = [[tweet.user.screen_name,tweet.user.location, tweet.text]
tweet_status = pd.DataFrame(data=status_text,
                             columns=['user','location','text'])
```

Gambar 2. Python data collection

Pada Gambar 3 ditampilkan contoh 10 baris data collection hasil dari streaming. Data collection hasil streaming data twitter di simpan di database kemudian dilakukan proses data preprocessing dengan cleaning data yang pertama menghapus url pada status tweet.

	user	location	text
0	gocornerblog	Ireland	When you think pubs/clubs will be able to open...
1	ChorusPhotog	Phoenixville, PA.	Small Town America: Day 41 / 2:20pm / April 28...
2	AgenciaAJN	Argentina	#Coronavirus. #Gaza reopens restaurants for #R...
3	cryptchart	AMS/RTH	3 tweets by @WHO that didn't age well. \n\n#wh...
4	NPPNewsTV	Pakistan	The #Coronavirus is Spreading Rapidly Around t...
5	Asylumseeker00		Final update: April 28, 11. 45 PM ISD\n\nTotal...
6	sboughorbel	Qatar	@MOPHQatar @SidraMedicine @QF @WHO @AJENEWS\n\...
7	LockdownNo	Earth	This clergyperson says #lockdown is "loving yo...
8	rtehrani	Trumbull, CT	Joseph D'Souza: Coronavirus shows how China's ...
9	BenSand54338716		So the news just told me a 10 year old will ru...

Gambar 3. Data collection COVID-19

Pada Gambar 4 diperlihatkan contoh data collection status tweet sebelum url di hilangkan. Tweet pada Gambar 4 memiliki beberapa elemen yang tidak diinginkan, URL dan tagar tidak akan dianalisis. Pada penelitian ini untuk menghapus URL (tautan) dan tagar menggunakan ekspresi reguler. Dalam fungsi yang didefinisikan dengan:

**re.sub** ('([<sup>0</sup>0-9A-Za-z \t])|(\ w +:\/\|\/\ S +) ',' '),

Fungsi **re.sub** digunakan untuk mengganti pilihan karakter yang ditentukan menggunakan ekspresi reguler, untuk menemukan semua string seperti URL dan tagar diganti dengan- "".

```
['A fully illustrated book to explain the corona virus to c
hildren and help parents to cope.\nhttps://t.co/P0mwhs8RBw...
https://t.co/cnKgv0NBpc', 'https://t.co/qitj1TKjk4\nI am h
onoured to present you all the NEW Rainbow Gaming Keyboard
and Mouse Set Multi-Colore... https://t.co/v7So92CvFA', '#Chi
na is threat for d world\n#EuropeanUnion #Europe #France #G
ermany #Italy #AngelaMerkel #COVID2019 #COVID@ 19... https://
t.co/fyv6QcZRK', 'Last updated: 2020-04-27 07:05:01+00:00G
MT\n Coronavirus [COVID-19]\n Cases: 2995757\n Death
s: 207021\n\n#corona... https://t.co/MVkpPsSaYmb', 'latest tim
es and details for #YomHazikaron/ #YomHaatzmout: #Israel #J
erusalem #corona #coronavirus #covid19 https://t.co/qDbezG2
vbT']
```

Gambar 4. Tweet preprocessing

Gambar 5 di tampilkan hasil cleaning data hapus url dan tagar. Kapitalisasi kata akan berpengaruh saat analisis text mining. Metode **string.lower()** digunakan untuk membuat setiap huruf capital menjadi kecil.

```
['A fully illustrated book to explain the corona virus to c  
hildren and help parents to cope', 'I am honoured to presen  
t you all the NEW Rainbow Gaming Keyboard and Mouse Set Mul  
tiColore', 'China is threat for d worldEuropeanUnion Europe  
France Germany Italy AngelaMerkel COVID2019 COVID19', 'Las  
t updated 20200427 0705010000GMT Coronavirus COVID19 Cases  
2995757 Deaths 207021corona', 'Latest times and details for  
YomHazikaron YomHaatzmaut Israel Jerusalem corona coronavi  
rus covid19']
```

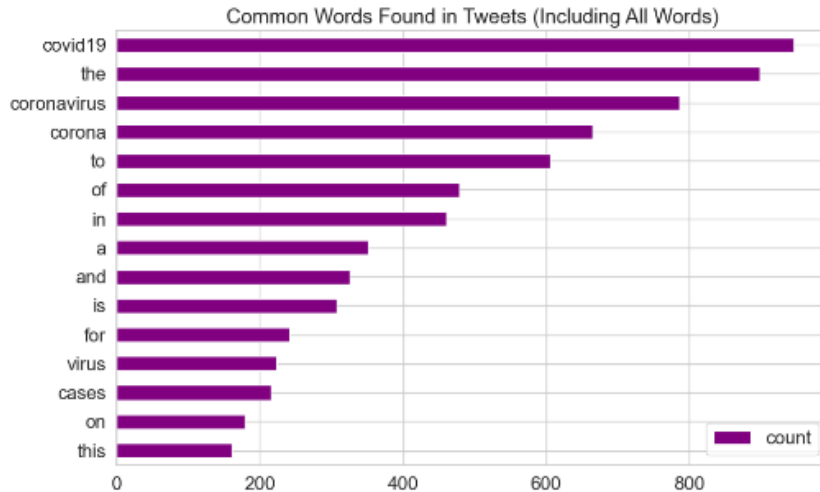
Gambar 5. *Tweet cleaning url*

Fungsi *set()* untuk menampilkan hanya kata-kata unik. Untuk melakukan analisis frekuensi kata, digunakan fungsi *split()* untuk memisah setiap kata menjadi elemen unik ditampilkan pada Gambar 6.

```
[['protocols', 'when', 'entering', 'your', 'home', 'during',  
'coronavirus', 'pandemicinvestmango', 'covid19', 'coronavirus  
'], ['live', 'we', 'have', 'mooted', 'to', 'continue', 'with',  
, 'the', 'lockdown', 'post', 'may', '3', 'with', 'relaxation',  
, 'on', 'activities', 'in', 'green', 'zones']]
```

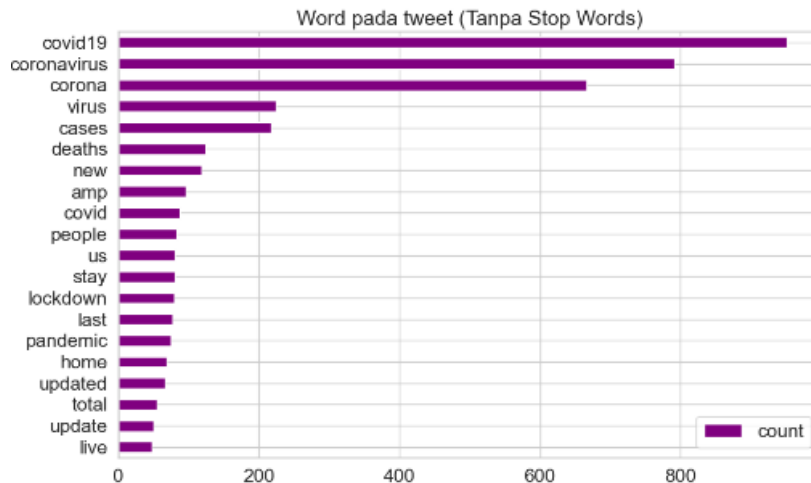
Gambar 6. *Tweet cleaning capital*

Untuk mendapatkan daftar *word frequency* berapa kali setiap kata muncul, menggunakan fungsi *most\_common* terdapat pada *Objek collection.Counter*, yang ditampilkan pada Gambar. 7



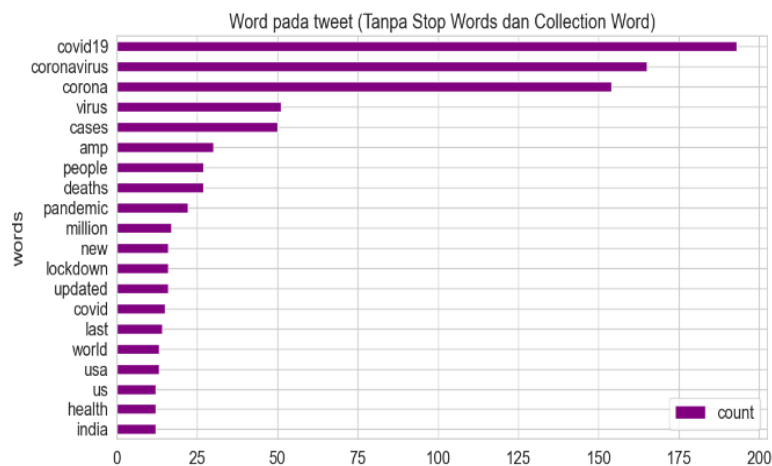
Gambar 7. *Tweet word frequency*

Bagian yang harus dihilangkan sebelum dilakukan analisis sentiment adalah *stopword*, kata-kata dengan informasi tidak bermakna untuk analisis *text mining*. *Library* Python *nlTK*, menyediakan daftar *stopword* yang dapat digunakan untuk membersihkan data Twitter. Fungsi *nlTK* hanya dapat bekerja pada semua kata dengan huruf kecil. Pada Gambar 8 di tampilkan *word frequency* dengan menghapus daftar *stopword*



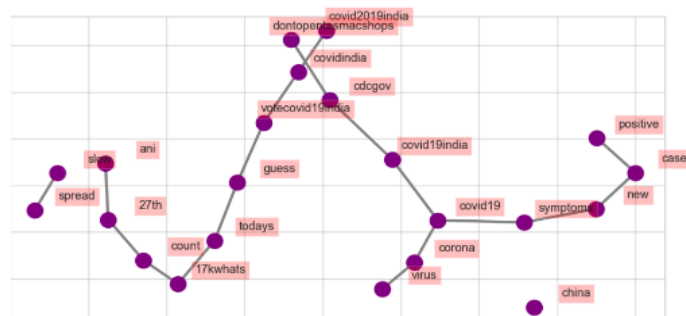
Gambar 8. *Tweet word frequency tanpa stopword*

Dengan menghapus *collection word* data yang di dapatkan dari data *cleaning* akan menjadi condong ke kata yang akan di analisa, yaitu variable COVID-19 atau coronavirus bahwa istilah ini akan ditemukan di setiap *tweet* . Pada Gambar 9 ditampilkan 20 daftar kata unik yang telah di proses data *cleaning* dari penghapusan *retweet* s, url, tagar, stopword dan data collection



Gambar 9. *Tweet data cleaning akhir*

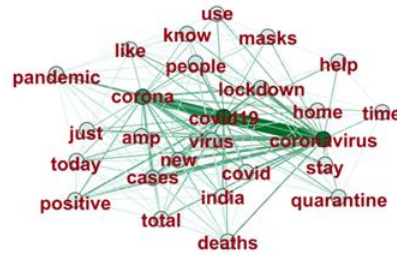
Untuk menganalisa hubungan antar kata yang memiliki kemunculan kata bersama di *tweet* , digunakan fungsi bigrams dari nltk. Daftar bigram kemunculan 20 kata di *tweet* pada Gambar 10.



Gambar 10. *Tweet bigram*



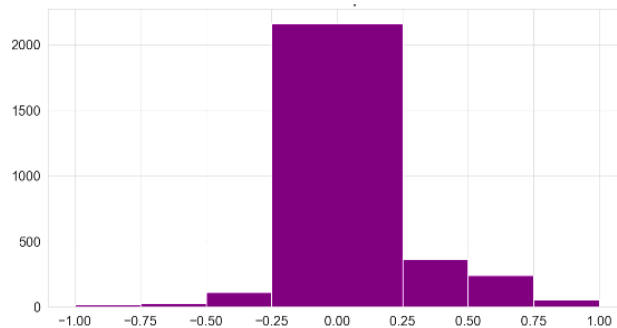
Hubungan antar kata ditampilkan pada Gambar 11 hasil analisa dengan fungsi *ngram* dengan *CountVectorizer*, dimana jumlah kemunculan kata dalam *tweet* di ubah menjadi matriks vektor .



Gambar 11. *Tweet* ngram

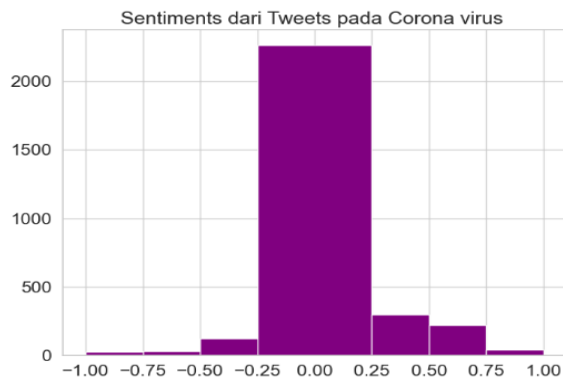
Berdasarkan data yang diperoleh, analisa hubungan antar kata *tweet* dengan ngram lebih fokus merujuk topik COVID-19 dibandingkan dengan bigram masih muncul daftar kata yang kurang merujuk ke topik COVID-19.

Untuk analisis sentimen dengan fungsi *Textblob* untuk mengidentifikasi opini kondisi pandemic COVID-19 dari data *tweet* . Nilai polaritas pada analisa sentiment ini antara 1 sampai -1. Opini dengan nilai polaritas mengarah ke nilai 1 menunjukkan opini dengan status positif, sedangkan polaritas mengarah ke nilai -1 menunjukkan negative, untuk opini netral nilai polaritas berkisar ke 0. Pada Gambar 12 ditampilkan grafik histogram dari nilai polaritas *tweet* COVID-19, nilai polaritas tertinggi pada status netral 58,94% yaitu polaritas status positif sebesar 33,68% dan negatif 7,36%



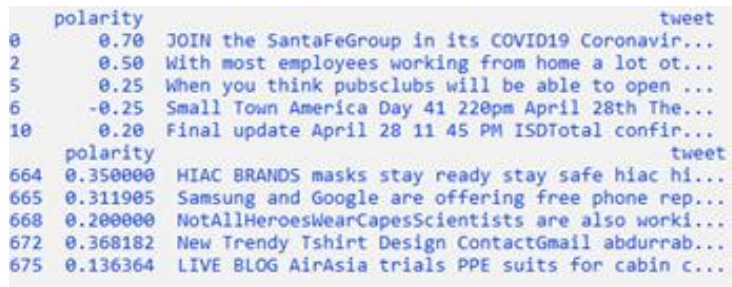
Gambar 12. Analisis sentimen COVID-19

Untuk nilai polaritas dengan variable pencarian coronavirus menunjukkan hasil analisa sentimen pada histogram Gambar 13, dengan jumlah nilai polaritas netral 55,10% positif 24,48% dan negatif 20,40%. Pada pencarian variable coronavirus lebih tinggi polaritas sentiment negative dari pada pencarian variable COVID-19.



Gambar 13. Analisis sentimen coronavirus

Pada Gambar 14 ditampilkan beberapa sampel status *tweet* dengan nilai polaritas tidak sama dengan 0 (!=0). Pada Tabel 1 menunjukkan beberapa sampel dari data collection *tweet* dari analisis sentiment variable #COVID-19 dan coronavirus dengan kategori positif, netral dan negatif serta nilai polaritasnya. Terdapat nilai polaritas -1 yang mengarah ke opini negatif tetapi terdapat *tweet* dengan polaritas 1 mengarah ke positif. Berdasarkan hasil analisa sentiment pada *tweet* yang mengarah pada polaritas negatif terdapat kata-kata yang dikategorikan sebagai opini negatif seperti *evil beer virus*, *terrible*, *worst*, sedangkan pada *tweet* positif terdapat kata-kata yang dikategorikan sebagai opini positif seperti *best*.



Gambar 14. *Tweet* polaritas (!=0)

Pada *tweet* polaritas sentimen netral tidak digunakan penggunaan kata-kata ungkapan perasaan opini.

Tabel 1. Kategori polaritas *tweet*

Nilai polaritas	<i>Tweet</i>	Kategori
-1	<i>I've got the mask hookup Click the link to protect yourself from the evil beer virus</i>	negatif
-1	<i>How you get infected with Covid, Corona in your office Terrible</i>	
-1	<i>Worst thing about Corona Introverts are no longer alone at home</i>	
0	<i>Coronavirus US officially surpasses one million cases coronavirus</i>	netral
0	<i>Kenyas COVID-19 case count increases to 374 14 people dead124 people recover</i>	
0	<i>Joseph DSouza Coronavirus shows how Chinas disregard for human rights threatens the world</i>	
1	<i>Best coronavirus news in a while scientist UniofOxford speeding ahead in developing COVID-19 vaccine</i>	positif
1	<i>Best facts to understandHighest VE Cases in USAMaximum Death in USAHighest</i>	
0.9	<i>Quarantine and Chill with your loved ones Grab the best shows to watch on the best laptops now</i>	

Semakin banyak penggunaan kata-kata ekspresi ungkapan perasaan opini baik yang mengarah ke sedih, marah, atau bahagia yang menyebabkan suatu kalimat atau *tweet* akan memiliki nilai polaritas ke arah negatif atau positif. Sedangkan *tweet* dengan polaritas sentiment netral memiliki presentase paling tinggi dibandingkan polaritas negative atau positif baik untuk pencarian variable COVID-19 atau coronavirus.

## V. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Pada penelitian ini dilakukan *streaming* data *tweet* dengan pencarian data collection dimulai dari 6 April 2020 dan dari hasil analisa sentimen baik pencarian variable COVID-19 atau coronavirus, nilai polaritas analisis sentimen kategori netral paling tinggi yaitu 58,94% untuk sentiment COVID-19 dan 55,10% untuk variabel coronavirus dibandingkan

polaritas negatif atau positif. Opini masyarakat pengguna twitter dengan status Bahasa Inggris tentang pandemi COVID-19 mulai bulan April 2020 melalui media sosial *tweet* masih dalam batas netral atau tidak banyak opini negatif yang mengarah ke kondisi kepanikan tetapi juga tidak banyak opini positif masyarakat dalam menghadapi kondisi pandemi COVID-19.

## 5.2 Saran

Pada penelitian ini, analisa sentiment *tweet* dengan python masih memiliki keterbatasan hanya dilakukan *streaming data tweet* dengan bahasa Inggris, pengembangan selanjutnya dapat dilakukan penelitian analisa sentiment dengan python untuk berbagai bahasa khususnya Indonesia. Data collection *tweet* yang disimpan masih belum besar dan banyak untuk mengarah ke bigdata.

## DAFTAR PUSTAKA

- Adawadkar, K. (2017). Python Programming-Applications and Future. *International Journal of Advance Engineering and Research Development*, 4(04), 1–4. <https://doi.org/10.21090/ijaerd.it032>
- Cheng, D., Schretlen, P., Kronenfeld, N., Bozowsky, N., & Wright, W. (2013). Tile based visual analytics for Twitter big data exploratory analysis. *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 2–4. <https://doi.org/10.1109/BigData.2013.6691787>
- Dhawan, Z. (2014). Big Data and Social Media. *Research Matters: A Cambridge Assessment Publication*, 18, 36–41.
- ECDC. (2020). Novel coronavirus disease 2019 ( COVID-19 ) pandemic: increased transmission in the EU / EEA and the UK – sixth update. *Rapid Risk Assessment*, 2019(March).
- Eroglu, Ç., Allen, N. J., Susman, M. W., Rourke, N. A. O., Young, C., Özkan, E., Chakraborty, C., Mulinyawe, S. B., Annis, D. S., Huberman, A. D., Green, E. M., Lawler, J., Dolmetsch, R., Christopher, K., Smith, S. J., Luo, Z. D., Rosenthal, A., Mosher, D. F., & Barres, A. (2018). *Intelligent Natural Language Processing: Trends and Applications*. 740(2), 380–392. <https://doi.org/10.1007/978-3-319-67056-0>
- Kiruthika, S. (2016). Sentiment Analysis of Twitter Data. *International Journal of Innovation in Engineering and Technology (IJJET)*, 6(4).
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. 2(11), 4704–4713.
- Martin, S., Karafillakis, E., Preet, R., & Wilder-Smith, A. (2020). The pandemic of social media panic travels faster than the COVID-19 outbreak Dengue Tools View project GBD 2013-2015 Study View project. *Article in Journal of Travel Medicine*, March. <https://doi.org/10.1093/jtm/taaa031/5775501>
- Yogapreethi, N. (2016). A Review on Text Mining in Data Mining. *International Journal on Soft Computing*, 7(2/3), 01–08. <https://doi.org/10.5121/ijsc.2016.7301>
- Praneeth Sai, J. V., & Balachander, B. (2020). Sentimental analysis of twitter data using tweepy and textblob. *International Journal of Advanced Science and Technology*, 29(3), 6537–6544.
- Saini, S. (2019). Sentiment Analysis on Twitter Data using R. *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 68–72.
- Setiowati, Y. (2016). Service Extraction and Sentiment Analysis to Indicate Hotel Service Quality in Yogyakarta based on User Opinion. *2018 International Seminar on*

*Research of Information Technology and Intelligent Systems (ISRITI)*, 427–432.

Sivarajah, U., Irani, Z., Gupta, S., & Mahroof, K. (2019). Role of big data and social media analytics for business to business sustainability: A participatory web context. *Industrial Marketing Management*, April, 0–1.  
<https://doi.org/10.1016/j.indmarman.2019.04.005>

Trupthi, M. (2017). *Sentiment analysis on twitter using streaming api*.  
<https://doi.org/10.1109/IACC.2017.177>