

Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19)

Implementasi Text Mining Untuk Deteksi Emosi Menggunakan Metode Leksikon (Studi Kasus: Twit Tentang Covid-19)

Agus Sasmito Aribowo¹, Siti Khomsah²

¹ Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

² Sains Data, Institut Teknologi Telkom Purwokerto, Indonesia

¹sasmito.skom@upnyk.ac.id, ^{2*}siti@ittelkom-pwt.ac.id

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: 1 December 2020

Revised: 18 January 2021

Accepted: 8 February 2021

Published: 28 February 2021

Abstract

Information and news about Covid-19 received various responses from social media users, including Twitter users. Changes in netizen opinion from time to time are interesting to analyze, especially about the patterns of public sentiment and emotions contained in these opinions. Sentiment and emotional conditions can illustrate the public's response to the Covid-19 pandemic in Indonesia. This research has two objectives, first to reveal the types of public emotions that emerged during the Covid-19 pandemic in Indonesia. Second, reveal the topics or words that appear most frequently in each emotion class. There are seven types of emotions to be detected, namely anger, fear, disgust, sadness, surprise, joy, and trust. The dataset used is Indonesian-language tweets, which were downloaded from April to August 2020. The method used for the extraction of emotional features is the lexicon-based method using the EmoLex dictionary. The result obtained is a monthly graph of public emotional conditions related to the Covid-19 pandemic in the dataset.

Abstrak

Keywords: Covid-19, Emotion Analysis, Lexicon
Kata kunci: Covid-19, Emosi, Leksikon

Informasi dan berita Covid-19 menuai ragam tanggapan masyarakat media sosial, salah satunya di Twitter. Perubahan opini netizen dari waktu ke waktu menarik untuk dianalisis, terutama pola sentimen dan emosi publik yang terkandung dalam opini tersebut. Kondisi sentimen dan emosi dapat menggambarkan tanggapan masyarakat terkait perkembangan pandemi Covid-19 di Indonesia. Penelitian ini mempunyai dua

tujuan yaitu, pertama mengungkap jenis emosi publik yang muncul selama pandemi Covid-19 di Indonesia. Kedua mengungkap topik atau kata yang paling sering muncul dalam setiap kelas emosi tersebut. Ada tujuh jenis emosi yang akan dideteksi yaitu marah (anger), takut (fear), jijik (disgust), sedih (sadness), terkejut (surprise), bahagia (joy), dan yakin (trust). Dataset yang digunakan adalah twit berbahasa Indonesia yang diunduh mulai bulan April hingga Agustus 2020. Metode yang akan digunakan untuk ekstraksi fitur emosi adalah metode lexicon based menggunakan kamus emosi EmoLex. Hasil yang diperoleh adalah grafik perkembangan kondisi emosi publik terkait pandemi Covid-19 setiap bulan dalam dataset.

1. Pendahuluan

Indonesia memasuki masa pandemi Covid-19 sejak awal Maret 2020. Informasi perkembangan Covid-19 baik resmi dari pemerintah maupun informasi dari pihak lain, banyak bertebaran di internet, terutama di media sosial. Pemberitaan terkait Covid-19 kerap memicu tanggapan bernada positif maupun negatif dari masyarakat. Setidaknya ada tiga media sosial utama yang paling sering digunakan masyarakat sebagai saluran penyampaian opini, yaitu Twitter, Youtube, dan Facebook. Pengguna media sosial (netizen) dapat memposting opini berupa teks, gambar, maupun video pada Facebook maupun Youtube. Emosi yang ditunjukkan oleh gambar dan video lebih mudah dikenali. Namun berbeda dengan kedua media tersebut, pengguna Twitter dibatasi hanya dapat memposting opini teks maksimal 140 karakter untuk sekali posting [1] dan ditingkatkan menjadi 280 pada September 2020. Ekstraksi informasi emosi melalui teks sepanjang 140 karakter tidak mudah dibandingkan dari teks yang panjang dan lengkap. Penggalan informasi berbasis teks ini masuk kedalam bidang *text mining*. Secara spesifik text mining yang membahas pengenalan sentimen atau polaritas pada komentar publik disebut sentiment analysis atau opinion mining.

Penelitian ini merupakan kelanjutan dari penelitian sebelumnya tentang deteksi kegelisahan masyarakat di tengah pandemi Covid-19 [2]. Menyebarnya informasi terkait Covid-19 mempengaruhi opini masyarakat di media sosial. Perubahan opini netizen dari waktu ke waktu menarik untuk dianalisis, terutama pola sentimen yang terkandung dalam opini tersebut. Sentimen menunjukkan ekspresi positif atau negatif. Liu (2012) mengatakan ekspresi negatif dapat menunjukkan sentimen kata (*word sentiment*) yaitu *bad*, *poor*, *terrible*. Sedangkan ekspresi positif dapat berupa *good*, *wonderfull*, dan *amazing*[3]. Chaudhuri (2006) dalam [3] mengatakan bahwa emosi sangat terkait dengan sentimen. Kekuatan sentimen sangat terhubung pada intensitas emosi. Misalnya, emosi 'sangat bahagia' menunjukkan sentimen sangat positif. Emosi sangat marah menunjukkan sentimen sangat negatif. Analisis emosi dalam teks, dari berita hingga postingan media sosial bermanfaat untuk mengetahui aspirasi emosi pengguna media sosial pada sebuah topik percakapan dan akan memahami bagaimana emosi membentuk perilaku manusia [4].

Secara teori psikologi, emosi terbagi menjadi enam kelas emosi dasar yaitu *surprise*, *happiness*, *anger*, *fear*, *disgust*, dan *sadness* [5]. Deteksi emosi adalah bidang kajian dalam *natural*

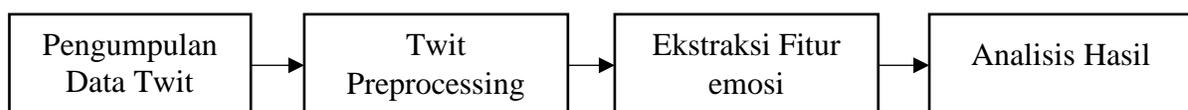
language processing (NLP). Aribowo *et al* (2020) menggunakan enam emosi dasar Ekman (1992) untuk mengkategorisasi fanatisme politik berdasarkan teks opini di Youtube [6]. Mohammad dan Turney (2013) menggunakan enam emosi dasar Ekman ditambah dua emosi *trust* dan *anticipation*, yang merupakan emosi dasar menurut Plutchik (1994). Delapan kelas emosi tersebut dipakai untuk membentuk kamus NRC Emotion Lexicon atau dikenal EmoLex [7]. EmoLex berisi daftar kata dan kaitannya dengan delapan emosi dasar yaitu *anger*, *anticipation*, *fear*, *surprise*, *sadness*, *joy*, *trust*, dan *disgust* serta dua sentimen yaitu negatif dan positif. Dengan demikian, EmoLex dapat mendeteksi sepuluh jenis emosi. Kamus EmoLex berisi 14.182 kosakata dan ditranslasi ke dalam banyak bahasa menggunakan Google Translate[8].

Lexical Based Approach adalah metode analisis sentimen dan emosi dengan menggunakan kamus sebagai bahasa atau sumber leksikal [9]. Bahkan metode leksikon pernah dipakai untuk deteksi *hate speech* [10]. Terkadang leksikon emosi dalam EmoLex tidak dapat mendeteksi jenis emosi yang terkandung dalam komentar atau opini. EmoLex dapat mengungkapkan emosi paling dominan yang terkandung dalam komentar Facebook komunitas diabetes [11], antara lain *joy*, *fear*, dan *sadness*. Leksikon emosi EmoLex sangat terbatas jika dibandingkan keragaman kata dalam bahasa Indonesia. Perluasan leksikon dapat dilakukan dengan pencarian sinonim leksikon EmoLex dengan Kateglo API [12]. Sinonimisasi cukup memperkaya kamus leksikon emosi, dibuktikan oleh penelitian Rohman *et al* (2019) yang dapat membuat kamus leksikon sampai dengan 34.872 kata. Namun, kamus leksikon tersebut saat digunakan untuk mendeteksi sepuluh jenis emosi, hanya mampu mendeteksi 55.45% dari 27.696 kata emosi yang diunduh dari komentar Facebook [12].

Penelitian mempunyai dua tujuan yaitu tujuan pertama mengungkap jenis emosi publik yang muncul dan perubahannya selama Pandemi Covid-19 di Indonesia. Tujuan kedua mengungkap topik atau kata yang paling sering muncul dalam setiap jenis emosi tersebut. Metode yang akan digunakan adalah *lexicon based* menggunakan EmoLex. Emosi yang akan diekstrak bersumber pada twit berbahasa Indonesia mulai bulan April hingga Agustus 2020. Ada tujuh emosi yang akan dideteksi yaitu marah (*anger*), takut (*fear*), jijik (*disgust*), sedih (*sadness*), terkejut (*surprise*), bahagis (*joy*), dan yakin (*trust*).

2. Metode

Penelitian berfokus pada perubahan emosi publik selama pandemi Covid-19. Subyek penelitian adalah masyarakat Indonesia. Populasi penelitian adalah pengguna twitter di Indonesia. Sampel yang digunakan adalah twit dalam bahasa Indonesia. Teknik pengambilan data twitter memanfaatkan API Twitter dan keyword terkait Covid-19. Rancangan alur penelitian ditunjukkan Gambar 1.



Gambar 1. Diagram alur penelitian

2.1. Pengumpulan Data Twitter

Pengambilan data dari website dikenal dengan istilah *scrapping*, tetapi tidak semua data website mudah diambil. Isu privasi data menjadi hambatan *scrapping* meskipun beberapa website justru memberikan fasilitas pengambilan data pada publik meskipun terbatas. Twitter menyediakan fasilitas Application Programming Interface (API) yang dapat digunakan semua user untuk mengunduh data *twit* [13]. Dengan keyword yang ditentukan, data *twit* yang mengandung keyword tersebut dapat diunduh.

Penelitian ini menggunakan Twitter API untuk mengumpulkan dataset berupa *twit* selama bulan April, Mei, Juni, Juli, dan Agustus tahun 2020. Keyword yang digunakan adalah *covid* dan *covid19*. Data yang diperoleh dan digunakan sebagai dataset penelitian dengan rincian jumlahnya per bulan ada pada Tabel 1.

Tabel 1. Jumlah *twit* sebagai dataset

Bulan-Tahun	Jumlah Opini
1 2020-04	7.210
2 2020-05	23.693
3 2020-06	12.771
4 2020-07	7.092
5 2020-08	2.192
6 2020-04	7.210
Total	42.675

2.2. *Twit* Preprocessing

Preprocessing adalah tahap paling penting dalam analisis teks. Data *twit* mengandung banyak *noise* seperti *hashtags*, *retweet*, *emoticon*, simbol, singkatan kata dan kata yang terkadang sulit dipahami. Seperti pada analisis sentimen dari data teks, secara umum pekerjaan preprocessing meliputi mengubah teks kedalam bentuk karakter seragam misalnya huruf kecil, menghilangkan tanda baca (*remove punctuation*), menghapus kata-kata tidak penting atau tidak dapat bermakna termasuk simbol *emoticon* (*stopword removal*).

Data *twit* mengandung *hashtags*, inipun tidak diperlukan dalam proses analisis. Sedangkan data re-*twit* pada setiap *twit* di proses sebagai data namun hubungan antara *twit* dan re-*twit* tidak dianalisis. Sehingga, data re-*twit* mempunyai mempunyai posisi sejajar dengan *twit*. Berikut contoh *raw data* *twit* pada Gambar 2.

RT:Bang...., udah dong bahas buzzernya...mari kita fokus bantu pemerintah Dan masyarakat akibat virus covid 19!.

Gambar 2. Contoh salah satu *twit*

Tahapan pre-processing *twit* antara lain *remove hashtags*, *remove unicode strings*, *lowercase*, *tokenizing*, *remove number*, *convert slangword*, dan *remove stopwords* [14], termasuk figur

atau nama orang/benda/subyek/obyek. Dalam *sentiment analysis* berbasis opini, konversi *slangword* dan *remove stopwords* adalah bagian paling penting. Kata *slangword* tidak akan dikenali oleh EmoLex sehingga harus dikonversi ke kata baku dalam bahasa Indonesia mengacu pada kosakata dalam kamus baku bahasa Indonesia (KKBI). Kamus konversi *slangword* ke dalam kata baku belum tersedia, sehingga kami harus mengembangkan sendiri. Dalam proses konversi *slangword*, penelitian ini menggunakan algoritma konversi pada penelitian sebelumnya[15]. Contoh setiap tahap proses tersebut pada Tabel 2 dan Tabel 3.

Tabel 2. Contoh hasil *pre-processing*

Kalimat asal Twit	Hasil Pre-processing			
	Remove Hashtags	Remove Unicode Strings dan Noise	Lowercase	Tokenizing
RT: Bang..., udah dong dong bahas buzzernya...mari kita fokus bantu pemerintah Dan masyarakat akibat virus covid 19!	Bang..., udah dong bahas buzzernya...mari kita fokus bantu pemerintah Dan masyarakat akibat virus covid 19!	Bang udah dong bahas buzzernya mari kita fokus bantu pemerintah Dan masyarakat akibat virus covid 19	bang udah dong bahas buzzernya mari kita fokus bantu pemerintah dan masyarakat akibat virus covid 19	"bang" "udah" "dong" "bahas" "buzzernya" "mari" "kita" "fokus" "bantu" "pemerintah" "dan" "masyarakat" "akibat" "virus" "covid" "19"

Tabel 3. Contoh hasil *pre-processing*

Hasil Tokenizing	Hasil Pre-processing		
	Remove number	Convert Slangword	Remove Stopword dan Figure
bang" "udah" "dong" "bahas" "buzzernya" "mari" "kita" "fokus" "bantu" "pemerintah" "dan" "masyarakat" "akibat" "virus" "covid" "19"	"bang" "udah" "dong" "bahas" "buzzernya" "mari" "kita" "fokus" "bantu" "pemerintah" "dan" "masyarakat" "akibat" "virus" "covid"	"bang" "sudah" "dong" "bahas" "buzzernya" "mari" "kita" "fokus" "bantu" "pemerintah" "dan" "masyarakat" "akibat" "virus" "covid"	"bahas" "buzzernya" "mari" "fokus" "bantu" "pemerintah" "masyarakat" "akibat" "virus" "covid"

Pada Tabel 3 terlihat bahwa hasil *convert slangword* hanya mengubah kata **udah** menjadi **sudah**. Sedangkan pada kolom hasil *remove stopwords* dan figur, kata **bang**, **dong**, dan **kita** terhapus. Hasil *stopword removal* dan *convert slangword* sangat tergantung pada keragaman kata yang digunakan sebagai penyaring *slangword* dan *stopword* tersebut.

2.3. Ekstraksi Fitur Emosi

Setiap kata dalam twit diasumsikan mengandung unsur-unsur emosi tertentu. Ekstraksi fitur dilakukan pada twit yang sudah bersih. Data twit yang telah menjadi token (kata tunggal) diproses untuk mendapatkan fitur emosi. Proses ekstraksi fitur tersebut menggunakan metode leksikon berbasis kamus emosi. Kamus emosi yang dipakai dalam penelitian ini adalah kamus Emotion Lexicon (EmoLex) [8]. Twit yang akan diolah dalam penelitian ini, selanjutnya disebut sebagai opini. Pembobotan emosi dikembangkan dari metode pembobotan pada penelitian

sebelumnya yaitu menggunakan rata-rata bobot [16]. Satu buah kata dalam opini bahkan bisa mengandung beberapa unsur emosi. Contohnya kata ‘unggul’ memiliki tiga unsur emosi yaitu senang, terkejut, dan yakin (percaya). Contoh lain kata ‘wabah’ memiliki unsur emosi ‘marah’, ‘jijik’, ‘takut’, ‘sedih’, dan ‘terkejut’. Jika D dataset opini, dan S adalah kalimat opini dalam dataset tersebut maka berlaku $D=[S1, S2, S3, \dots, Sn]$ dimana n adalah jumlah opini. Jika W adalah kata dalam setiap kalimat opini S maka berlaku $S=[W1, W2,..Wm]$ dimana m adalah jumlah kata dalam kalimat opini S . Algoritma ekstraksi fitur emosi mengikuti alur algoritma berikut.

- a. Mulai
- b. Emosi $E=[anger, fear, surprise, sadness, joy, trust, and disgust]=[E1,E2,..Ex]$. Jumlah emosi adalah X dimana $X=7$
- c. Baca D dimana $D=[S1, S2, S3, \dots, Sn]$ dan $n=$ jumlah opini
- d. Tentukan $nextbaris = 1$
- e. Kerjakan sebanyak n :
 - 1) Baca teks opini $S[nextbaris]$
 - 2) Parsing teks opini S menjadi sekumpulan kata yang mandiri W , sehingga $S=[W1, W2,..Wm]$ dimana m adalah jumlah kata dalam opini $S[nextbaris]$
 - 3) Cek setiap kata mandiri W tersebut emosi yang berlaku di kamus $NRC (EmoLex)$
 - 4) Catat jumlah setiap emosi dalam opini tersebut sebagai score bagi setiap jenis emosi.
 - 5) Pilih emosi yang paling maksimal nilai skor-nya.
 - 6) Untuk menentukan emosi yang kuat dan lemah maka skor setiap emosi dibagi dengan skor emosi yang paling maksimal.
 - a) Jika hasil pembagiannya di atas 0 maka emosi tersebut diberi label bobot 1.
 - b) Jika hasil pembagiannya 0 maka emosi tersebut diberi label bobot 0.
 - 7) Tentukan $nextbaris++$
- f. Jika $nextbaris \leq$ jumlahrecord maka Kembali ke poin e .
- g. Finish.

Contoh ekstraksi fitur emosi based EmoLex pada Tabel 4.

Tabel 4. Contoh hasil ekstraksi emosi menggunakan EmoLex

No	Opini (twit)	Jenis emosi						
		Marah	Takut	Jijik	Sedih	Terkejut	Bahagia	Yakin
0	Seharusnya saat ini...	0	1	0	1	0	0	0
1	Klo kata gw tuh lagi menerapkan Her...	1	1	0	1	0	1	1
2	Saat ini yang bisa saya lakukan hanya menyiapk...	0	0	0	0	0	1	1
3	Satu Warga Positif Corona, Bupati: Kondisi ini...	1	1	0	1	0	0	0

Angka 1 pada kolom jenis emosi menunjukkan bahwa emosi tersebut eksis dalam opini yang terkait dengannya. Selanjutnya, untuk mengetahui bobot emosi setiap hari maka data opini yang sudah berlabel untuk setiap jenis emosi dilakukan proses perhitungan rata-rata hariannya. Hasil akhir proses rata-rata setiap emosi perhari ditunjukkan Tabel 5.

Tabel 5. Contoh hasil perhitungan rata-rata score emosi setiap hari

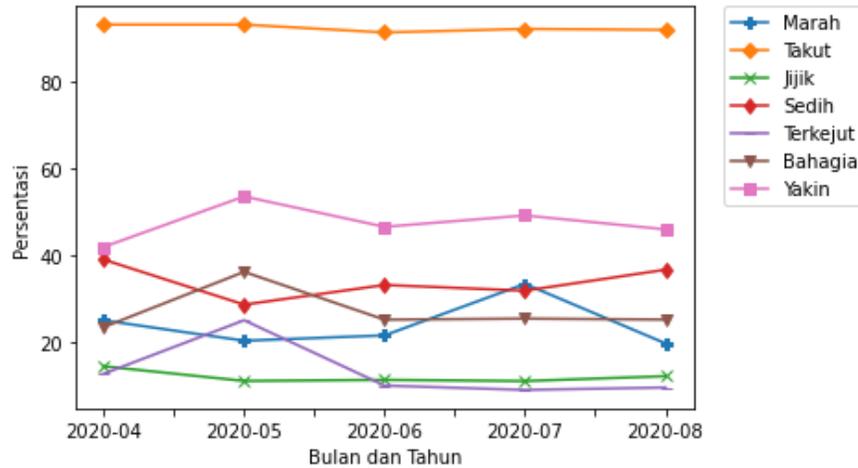
Date	Marah	Takut	Jijik	Sedih	Terkejut	Bahagia	Yakin
2020-04-21	0.353276	0.957265	0.198481	0.485280	0.210826	0.345679	0.130104
2020-04-22	0.255245	0.946387	0.167832	0.414918	0.248252	0.407925	0.120047
2020-04-23	0.247002	0.929257	0.107914	0.365707	0.175060	0.359712	0.081535
2020-04-24	0.208606	0.933583	0.140318	0.385407	0.190833	0.423761	0.093545
2020-04-25	0.237323	0.925963	0.135903	0.360041	0.255578	0.403651	0.088235
2020-04-26	0.274419	0.949612	0.167442	0.373643	0.201550	0.368217	0.099225
2020-04-27	0.229464	0.940179	0.107143	0.354464	0.191964	0.422321	0.103571

3. Hasil dan Pembahasan

Berdasarkan proses ekstraksi emosi, diperoleh hasil rata-rata nilai bobot emosi mulai bulan April sampai dengan Agustus 2020, pada Tabel 6. Sedangkan perubahan emosi pada bulan-bulan tersebut ditunjukkan oleh grafik pada Gambar 3.

Tabel 6. Hasil perhitungan rata-rata score emosi setiap bulan

Bulan – Tahun	Marah	Takut	Jijik	Sedih	Terkejut	Bahagia	Yakin
2020-04	0.258669	0.940915	0.147434	0.390846	0.102497	0.209709	0.390153
2020-05	0.211708	0.942177	0.115013	0.285696	0.225974	0.344912	0.512261
2020-06	0.225354	0.923420	0.116749	0.329497	0.072900	0.228408	0.440138
2020-07	0.347857	0.935279	0.123096	0.326142	0.079808	0.238156	0.470812
2020-08	0.207573	0.927464	0.123175	0.361770	0.068431	0.229471	0.436131



Gambar 3. Pergerakan emosi publik terhadap Pandemi Covid-19 dari April – Agustus 2020

Gambar 3 menunjukkan bahwa ada emosi paling kuat yaitu rasa takut, yang keberadaanya terus ada selama bulan April sampai dengan Agustus 2020. Ada satu jenis emosi yang dominan daripada yang lain yaitu takut. Jenis emosi berikutnya adalah yakin. Jenis-jenis emosi yang lain seimbang dalam jumlah terutama emosi jijik (disgust). Emosi marah, bahagia dan sedih cenderung silih berganti dalam setiap bulan. Khusus emosi marah meningkat cukup signifikan pada bulan Juli, tapi menurun drastis di bulan Agustus. Emosi bahagia naik signifikan di bulan Mei, tetapi menurun dan stabil di bulan-bulan berikutnya. Emosi sedih tinggi di bulan April, dan menurun stabil di bulan-bulan berikutnya. Kondisi emosi tersebut dapat diketahui dengan membaca keyword yang dominan dibahas pada setiap bulan. Tabel 7 menunjukkan beberapa kata kunci topik pembicaraan muncul pada setiap jenis emosi.

Tabel 7. Beberapa term (kata) utama dalam setiap kelas emosi

No	Terkejut		Bahagia		Yakin	
	Kata	N	Kata	N	Kata	N
1	covid	9833	Covid	17722	Covid	29023
2	pemerintah	9508	pemerintah	16561	pemerintah	27348
3	Tidak mudik	5043	tidak mudik	5912	Tidak mudik	6817
4	Memutus mata rantai	3793	Memutus mata rantai	4071	kebijakan	5073
5	kebaikan	3680	kebaikan	3851	Memutus mata rantai	4476
6	imbau	3632	imbau	3776	masyarakat	4464
7	polri	2426	polri	2748	Polri	4449
8	stopmudikcegahcovid	2317	kesehatan	2726	penyebaran	4303

Tabel 8. Lanjutan

NO	Marah		Takut		Jijik		Sedih	
	Kata	N	Kata	N	Kata	N	Kata	N
1	covid	13863	covid	54986	covid	7224	covid	18751
2	pemerintah	13129	pemerintah	52264	pemerintah	6766	pemerintah	18402
3	masyarakat	2682	pandemi	9350	wabah	1385	pandemi	6830
4	pandemi	2258	masyarakat	7786	masyarakat	955	masyarakat	2409
5	kebijakan	2222	tidak mudik	7375	pandemi	836	penanganan	1872
6	penyebaran	1943	penyebaran	6941	kebijakan	651	orang	1535
7	kesehatan	1494	penanganan	5778	corona	638	virus	1522
8	virus	1305	kebijakan	5650	virus	635	normal	1522
9	protokol	1223	virus	5543	orang	608	wabah	1515

Maka wordcloud dapat dipakai untuk memantau keyword dominan dari percakapan setiap bulan tersebut. Gambar 4 menunjukkan topik pada bulan April ketika Indonesia memasuki masa pandemi untuk bulan kedua. Pada bulan ini pemerintah banyak mengeluarkan kebijakan untuk menangani pandemi covid-19. Dampak Covid-19 semakin terlihat dengan banyaknya korban dan persebaran virus yang cukup masif. Banyak berita terkait tema PSBB, lockdown cukup gencar diperbincangkan di media sosial.

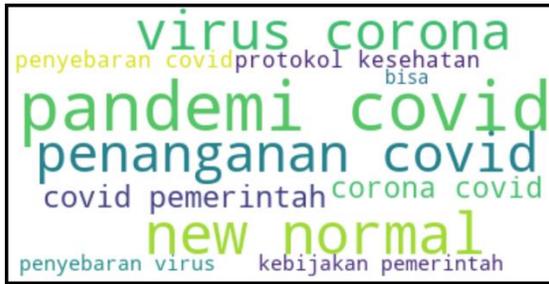


Gambar 3. Wordcloud percakapan umum di twitter pada April 2020



Gambar 4. Wordcloud percakapan umum di twitter pada Mei 2020

Emosi yang dominan pada bulan ini masih relatif sama dengan bulan sebelumnya yaitu takut. Takut sangat dimungkinkan karena pandemi masuk ke masa puncak menurut para pakar. Gambar 5 menunjukkan tema percakapan pada bulan Mei. Emosi yang diekstraksi dari percakapan di media sosial ada yang mengalami peningkatan cukup signifikan, yaitu bahagia. Emosi tersebut dapat disebabkan oleh kondisi bulan mei yang merupakan bulan puasa dan lebaran idul fitri dimana masyarakat sedikit melupakan pandemi, tetapi tetap menerapkan himbauan pemerintah. Pada bulan ini berdasarkan *wordcloud* adalah berfokus pada himbauan pemerintah untuk memutus mata rantai Covid-19. Himbauan untuk tidak mudik, untuk memutus mata rantai covid-19 dan mencegah penularan ke daerah menjadi topik utama percakapan di media sosial. Peran serta Polri untuk memberikan himbauan dan menyukseskan program pemerintah juga menjadi pembicaraan utama, terutama terkait larangan untuk nekad mudik dan adanya sanksi.



Gambar 6. Wordcloud percakapan umum di twitter pada Juni 2020



Gambar 7. Wordcloud percakapan umum di twitter pada Juli 2020

Pada bulan Juni, masyarakat banyak membicarakan hal baru yaitu tentang *new normal*. Gambar 6 menunjukkan tema percakapan tentang penanganan pandemi Covid-19 oleh pemerintah masih menjadi topik utama. Penyebaran virus dan penerapan protokol kesehatan sebagai salah satu cara dalam menangani penyebaran Covid-19 tetap digaungkan. Gambaran percakapan menunjukkan emosi sedih meningkat cukup signifikan menggantikan emosi bahagia. Hal ini bisa jadi disebabkan mulai adanya kejenuhan akibat adanya PSBB, *physical distancing*, dan kesulitan ekonomi. Gambar 7 memperlihatkan topik percakapan bulan Juli masih sama dengan percakapan bulan Juni yaitu banyaknya percakapan tentang penerapan protokol Kesehatan antara lain *physical distancing*, penggunaan masker untuk pencegahan penyebaran Covid-19.



Gambar 5. Wordcloud percakapan umum di twitter pada Agustus 2020

Banyaknya percakapan masih terkait protokol kesehatan dan mulai banyak percakapan tentang pemulihan ekonomi nasional. Ketakutan yang ada mungkin mulai berkurang, terlihat dari jenis emosi marah mengalami penurunan drastis seperti pada grafik Gambar 3.

4. Kesimpulan

Penelitian ini bertujuan mendeteksi jenis emosi dalam percakapan Twitter selama pandemi Covid-19 sejak April – Agustus 2020. Berikutnya, kata-kata apa saja yang paling sering muncul dalam setiap jenis emosi tersebut. Ekstraksi menggunakan EmoLex dari 42 ribu lebih twit yang diperoleh sepanjang April sampai dengan Agustus 2020, menunjukkan yang paling dominan dan terus ada dalam percakapan adalah emosi takut sebesar 92% sampai dengan 94%. Berikutnya, emosi yakin konsisten menempati urutan kedua dengan nilai 39% sampai dengan 51%. Meskipun emosi sedih konsisten keberadaannya dari April sampai Agustus, namun persentasinya rendah, 28 % sampai 39% saja. Begitu juga dengan emosi bahagia cenderung rendah antara 20% sampai 34%. Emosi terkejut sempat naik di bulan Mei yaitu pada presentasi

10%, namun justru terus menurun sampai Agustus hingga 0,06% saja. Meskipun presentasi percakapan dengan emosi marah relatif kecil, namun sempat naik pada bulan Juli. Sedangkan emosi jijik, dari April sampai Agustus 2020 tidak tinggi.

Daftar Pustaka

- [1] M. Lailiyah, S. Sumpeno, and I. K. E. Purnama, "Sentiment analysis of public complaints using lexical resources between Indonesian sentiment lexicon and sentiwordnet," in *2017 International Seminar on Intelligent Technology and Its Application: Strengthening the Link Between University Research and Industry to Support ASEAN Energy Sector, ISITIA 2017 - Proceeding*, 2017, vol. 2017-Janua, pp. 307–312, doi: 10.1109/ISITIA.2017.8124100.
- [2] Y. Fauziah, S. Saifullah, and A. S. Aribowo, "Design Text Mining for Anxiety Detection using Machine Learning based-on Social Media Data during COVID-19 pandemic," *Proceeding LPPM UPN "Veteran" Yogyakarta Conf. Ser. 2020–Engineering Sci. Ser.*, vol. 1, no. 1, pp. 253–261, 2020, doi: 10.31098/ess.v1i1.117.
- [3] B. Liu, *Sentiment Analysis and Opinion Mining*. 2012.
- [4] S. Mohammad, "Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text," *Emot. Meas.*, pp. 201–237, 2021, doi: 10.1016/B978-0-08-100508-8.00009-6.
- [5] P. Ekman, "An argument for basic emotions An Argument for Basic Emotions," no. October 2014, pp. 37–41, 2008, doi: 10.1080/02699939208411068.
- [6] A. S. Aribowo, H. Basiron, N. S. Herman, and S. Khomsah, "Fanaticism Category Generation Using Tree-Based Machine Learning Method Fanaticism Category Generation Using Tree-Based Machine Learning Method," *J. Phys. Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1501/1/012021.
- [7] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word–Emotion Association Lexicon," no. 2010, pp. 1–25, 2013.
- [8] S. M. Mohammad, "Sentiment and Emotion Lexicons." .
- [9] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/715730.
- [10] N. D. Gitari, Z. Zuping, H. Damien, and J. Long, "A Lexicon-based Approach for Hate Speech Detection," *Int. J. Multimed. Ubiquitous Eng.*, vol. 10, no. 4, pp. 215–230, 2015, doi: 10.14257/ijmue.2015.10.4.21.
- [11] V. Balakrishnan, M. C. Martin, W. Kaur, and A. Javed, "A comparative analysis of detection mechanisms for emotion detection," 2019, doi: 10.1088/1742-6596/1339/1/012016.
- [12] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," pp. 70–76, 2019, doi:

10.30864/eksplora.v9i1.277.

- [13] “<https://developer.twitter.com/en/docs>.” .
- [14] S. Mujilawati, “Pre-Processing Text Mining Pada Data Twitter,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 2089–9815, 2016.
- [15] S. Khomsah and A. S. Aribowo, “Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia,” *Rekayasa Sist. dan Teknol. Informasi, RESTI*, vol. 4, no. 10, pp. 648–654, 2020, doi: <https://doi.org/10.29207/resti.v4i4.2035>.
- [16] A. S. Aribowo, “Arsitektur Aplikasi Twitter Opinion Mining Untuk Mengetahui Sentimen Publik Terhadap Merek,” in *Seminar Nasional Informatika (semnasIF) 2015 UPN Veteran Yogyakarta*, 2015, vol. 2015, no. November, pp. 14–20.