

ANALISIS SENTIMEN TAGAR #INDONESIA TERSEERAH DI MASA COVID-19 MENGUNAKAN METODE *SENTISTRENGTH*

Bisma Aulia¹, Pradita Eko Prasetyo Utomo², Ulfa Khaira³ dan Tri Suratno⁴

^{1,2,3,4}Program Studi Sistem Informasi, Universitas Jambi, Jl. Lintas Sumatera Jl. Jambi – Muara
Bulian Km.15, Mendalo Darat, Muara Jambi, Jambi, Indonesia

¹Email: bismaaulia82@gmail.com

²Email: pradita.eko@unja.ac.id

³Email: ulfa.ilkom@gmail.com

⁴Email: trie123@gmail.com

ABSTRAK

Penggunaan sosial media di kala pandemi saat ini menjadi pilihan masyarakat dalam mengekspresikan pikirannya, salah satunya adalah Twitter. Dengan adanya fitur tagar di aplikasi Twitter, masyarakat dapat mengetahui informasi terbaru yang sedang tren. Dengan kondisi pandemi saat ini yang memunculkan banyak masalah sosial, politik, ekonomi dan sebagainya, membuat Twitter menjadi tempat masyarakat mengekspresikan emosinya. Belum lama ini, tagar #IndonesiaTerseerah menjadi buah bibir di masyarakat karena menggambarkan kekecewaan masyarakat terhadap penanganan virus Corona (COVID-19) di Indonesia. Penelitian ini bertujuan melihat bagaimana sentimen masyarakat Indonesia melalui tagar #IndonesiaTerseerah. Sentimen tersebut dianalisis melalui algoritma *sentistrength*, dan diklasifikasikan menjadi 3 kelas, yakni positif, netral, dan negatif. Algoritma ini menggunakan dasar leksikon sebagai penghitungan bobot kekuatan sentimennya. Tahapan yang dilakukan dalam penelitian ini yakni tahap *crawling* data, *preprocessing* data dan pembobotan kata. Hasil dari penelitian ini didapatkan 236 *tweet* data dengan 41,5% bersentimen netral, 32,2% bersentimen negatif, dan 26,3% bersentimen positif. Penelitian ini diharapkan dapat menjadi tolak ukur pemangku kepentingan dalam mengambil sebuah keputusan.

Kata kunci: #IndonesiaTerseerah, COVID-19, *sentistrength*, Twitter, analisis sentimen

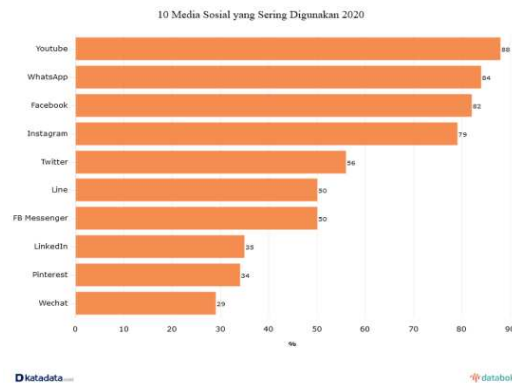
ABSTRACT

The use of social media during the current pandemic is the choice of the community in expressing their thoughts, one of which is Twitter. With the hashtag feature in the Twitter application, people can find out the latest trending information. With the current pandemic condition that raises many social, political, economic problems and so on, making Twitter a place for people to express their emotions. Not long ago, the hashtag #IndonesiaTerseerah became a byword in the community because it described the public's disappointment with the handling of the Corona virus (COVID-19) in Indonesia. This study aims to see how the sentiments of the Indonesian people through the hashtag #IndonesiaTerseerah. The sentiments were analyzed through the sent-strength algorithm, and classified into 3 classes, namely positive, neutral, and negative. This algorithm uses the lexicon as the basis for calculating the weight of the sentiment strength. The stages carried out in this study are the data crawling stage, data preprocessing and word weighting. The results of this study obtained 236 tweet data with 41.5% neutral sentiment, 32.2% negative sentiment, and 26.3% positive sentiment. This research is expected to be a benchmark for stakeholders in making a decision.

Keywords: #IndonesiaTerseerah, COVID-19, *sentistrength*, Twitter, sentiment analysis

1. PENDAHULUAN

Meningkatnya kasus COVID-19 di Indonesia membuat masyarakat bekerja, belajar, dan bermain dari rumah. Akibatnya terjadi kenaikan penggunaan sosial media saat ini. Penggunaan media sosial yang melonjak di tengah COVID-19, penggunaan Youtube sebagai media yang paling banyak diakses melonjak 98%, diikuti oleh WhatsApp, Facebook, Instagram, serta Twitter [1]. Penggunaan media sosial tersebut dijadikan wadah penyalur kreativitas, hobi, ataupun sebagai media kritik dan berkeluh kesah masyarakat. Gambar 1 menunjukkan media sosial yang sering digunakan selama tahun 2020.



Gambar 1. Media Sosial yang Sering digunakan 2020 [2].

Seperti yang dilihat dari *traffic* penggunaan media sosial di Indonesia pada 2020, Twitter merupakan salah satu media sosial yang masuk ke jajaran media yang paling banyak digunakan sebanyak 56%. Penggunaan Twitter menjadi pilihan bagi warganet untuk menuangkan keluh kesah, cerita, hingga pengalaman pribadinya. Fitur tagar dalam penggunaan Twitter menjadi salah satu keunggulan Twitter. Dengan adanya fitur ini masyarakat bisa dengan mudah mengetahui berita dan informasi terbaru yang sedang tren. Beberapa tagar yang tren di kalangan masyarakat sebagian berhubungan dengan COVID-19.

Kehadiran Twitter selain menjadi media hiburan bagi masyarakat juga dapat menjadi media dalam mengetahui sentimen masyarakat terhadap suatu kebijakan, berita, atau informasi, seperti kebijakan kenaikan harga premi jaminan sosial yang hangat dibahas pada tahun 2019. Kini, tren yang berkaitan dengan COVID-19 seperti tagar *#IndonesiaTerserah* menjadi populer terlebih pada 14 November 2020. Tagar ini diserukan oleh *influencer* serta relawan COVID-19, dr.Tirta Mandira. Tagar ini hangat dibahas oleh masyarakat karena keresahan masyarakat terhadap penanganan virus Corona di Indonesia. Meski bagi sebagian pihak, Twitter hanya sebagai media hiburan, tetapi dari penjelasan tersebut, Twitter dapat dijadikan tolak ukur oleh pemangku kepentingan dalam mengetahui pendapat masyarakat terhadap topik tertentu. Pendapat tersebut akan diklasifikasikan menjadi positif, negatif, atau netral melalui pengolahan data teks.

Pengolahan data teks yang sering disebut *text mining* adalah proses mengekstraksi sebuah data yang berupa teks yang tidak terstruktur yang dibantu perangkat lunak yang dapat mengidentifikasi pola dalam data untuk mengambil sebuah keputusan [3]. Dalam penentuan perasaan yang terkandung dalam teks tersebut digunakan analisis sentimen atau *opinion mining* yang merupakan metode untuk mengekstrak data opini, data tekstual secara otomatis untuk melihat sentimen yang terkandung [4].

Penelitian [5] melakukan analisis sentimen Twitter menggunakan metode *deep belief network*. Penelitian [6] melakukan sentimen analisis *cyber bullying* pada media sosial Twitter dengan algoritma *sentiStrength* dan penelitian ini mendapatkan nilai akurasi sebesar 60,5%.

Pemilihan metode *SentiStrength* sebagai penelitian karena pada penelitian sebelumnya akurasi yang dilakukan masih rendah serta ada beberapa tahap yang tidak dilakukan, serta masih sedikitnya penggunaan metode ini dalam Bahasa Indonesia.

Tujuan penelitian ini untuk mengetahui bagaimana perasaan masyarakat atas keresahan mereka melalui tagar *#IndonesiaTerserah*. Penelitian ini bermanfaat bagi pemangku kepentingan untuk mengambil keputusan terkait keresahan masyarakat akan kenaikan kasus COVID-19 di Indonesia.

2. MATERI DAN METODE

Penelitian ini menggunakan bebantuan aplikasi Anaconda dengan bahasa pemrograman Python versi 3.8.3 dan menggunakan IDE Jupyter Notebook 6.0.3. Tahapan dalam penelitian ini adalah pengumpulan data (*crawling data*), pemrosesan awal data, ekstraksi fitur, pemodelan, dan evaluasi, seperti pada gambar 2.

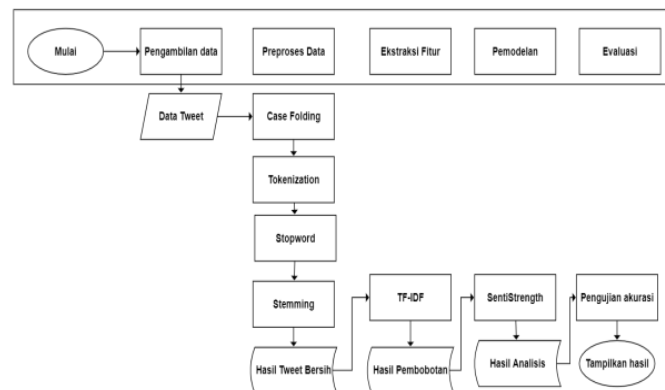
Crawling Data

Crawling data adalah tahap mengumpulkan data dari *database* [7]. Dalam penelitian ini, dataset yang digunakan merupakan data *tweet* dari *crawling* Twitter melalui API Twitter dengan kata kunci *#IndonesiaTerserah* sebanyak 3651 *tweet*. Proses *crawling* data dilakukan hingga 24 Desember 2020.

Pemrosesan Awal Data

Pemrosesan awal data bertujuan untuk membersihkan data *tweet* yang akan digunakan. Pada tahap ini dilakukan *case folding* (mengubah huruf besar menjadi huruf kecil), menghapus *mention* (@), menghapus tautan, menghapus karakter-karakter tertentu (misalnya ~:), *hashtag* (#), *retweet* (RT),

menghapus angka, menghapus *stopword* (kata yang tidak perlu), menghapus duplikasi, proses *tokenization* (membagi kalimat perkata), dan proses *stemming* (menghapus imbuhan). Pada proses ini, *tweet* yang tersisa sebanyak 241 *tweet*.



Gambar 2. Alur Penelitian

Ekstraksi Fitur

Komputer tidak dapat mengolah data selain dari numerik, maka diperlukan proses untuk mengekstrak data non-numerik menjadi numerik, proses pengekstrakan data ini disebut dengan ekstraksi fitur. Dalam penelitian ini, digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF adalah perhitungan yang menggambarkan seberapa penting sebuah kata dalam sebuah dokumen [8]. TF-IDF memiliki perhitungan seperti pada persamaan 1.

$$W_{dt} = tf_{dt} * IDF_t \dots\dots\dots (1)$$

di mana:

- W : bobot dokumen ke-d terhadap kata ke-t
- d : dokumen ke-d
- t : kata ke-t dari kata kunci
- tf : banyaknya kata yang dicari pada sebuah dokumen
- IDF : nilai IDF didapatkan melalui rumus $IDF = \log_2(D/df)$
- D : total dokumen

Pemodelan

Pemodelan algoritma yang dilakukan dalam menganalisis sentimen adalah algoritma *sentistrength*. Algoritma ini dirancang untuk *text web* untuk memprediksi kekuatan sentimen positif dan negatif teks secara bersamaan menggunakan pendekatan leksikon [9]. Algoritma *sentistrength* dapat berjalan dalam dua model yakni *supervised* dan *unsupervised* [9]. Dalam penelitian ini, peneliti menggunakan model *unsupervised* dalam prosesnya. *Unsupervised* adalah pendekatan yang tidak menggunakan data latih dalam melakukan prediksi atau klasifikasi. Algoritma *SentiStrength* adalah:

If positive value > negative value then
positive sentiment
If positive value < negative value then
negative sentiment
If positive value = negative value then **neutral**
sentiment

Rentang nilai dari algoritma ini dimulai dari nilai 1 sampai 5. Nilai 1 menunjukkan kalimat tidak bersentimen positif atau negatif, sedangkan nilai 5 menunjukkan sentiment yang bernilai positif atau negatif [6].

Evaluasi

Pada tahap ini dilakukan pengujian akurasi dari hasil yang telah didapatkan dari proses sebelumnya. Tahap ini dilakukan perbandingan antara data aktual dan data prediksi. Data aktual adalah data yang dihasilkan oleh pelabelan manual, sedangkan data prediksi adalah pelabelan yang dihasilkan oleh sistem. Penelitian ini dilakukan uji akurasi dengan membandingkan kasus yang terklarifikasi benar dengan jumlah semua kasus, metode yang digunakan dalam uji akurasi ini adalah *confusion matrix*.

Confusion matrix adalah pengukuran kinerja untuk masalah klasifikasi *machine learning* di mana keluaran atau *output* dapat berupa 2 kelas atau lebih [10]. *Confusion matrix* ditampilkan pada tabel 1 [6].

Tabel 1. *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

Perhitungan dari akurasi dari *confusion matrix* seperti pada persamaan 2.

$$Akurasi = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \dots\dots\dots (2)$$

3. HASIL DAN PEMBAHASAN

Crawling Data

Data dalam penelitian ini diambil dari media sosial Twitter dengan keyword #IndonesiaTerserah. Data yang terkumpul sebanyak 3651 *tweet*. Data ini belum dapat dianalisis sentimen karena perlu dilakukan pembersihan data. Hasil pengumpulan data dapat dilihat pada tabel 2.

Tabel 2. Hasil pengumpulan data Twitter

Teks
Indonesia hebat, kaya, dan besar hanya saja kebanyakan manusianya lemah, miskin, dan kecil dalam berperilaku #indonesiaterserah
RT @tirta_hudhi: #indonesiaterserah
Ini bukan soal cebong kampret
Ini soal kemanusiaan
Ini soal keadilan
Ini soal pengorbanan
@detikcom Bakalan bikin klaster baru nih #indonesiaterserah

Pemrosesan Awal Data

Tahap preproses data adalah tahap yang perlu dilakukan agar hasil analisis bersih dari duplikasi data, tanda baca (titik, koma, tanda tanya, tanda seru, dan sebagainya), simbol @, *Retweet* (RT) dan sebagainya. Selain itu, pada proses ini akan dilakukan proses *case folding*, penghapusan *stop word*, *tokenization*, serta *stemming*. *Case folding* adalah mengonversi teks dalam dokumen menjadi *lowercase*. Hasil dari perubahan teks menjadi *lowercase* terdapat pada tabel 3.

Tabel 3. Hasil proses *case folding*

Teks
indonesia hebat, kaya, dan besar hanya saja kebanyakan manusianya lemah, miskin, dan kecil dalam berperilaku #indonesiaterserah
rt @tirta_hudhi: #indonesiaterserah
ini bukan soal cebong kampret
ini soal kemanusiaan
ini soal keadilan
ini soal pengorbanan
@detikcom bakalan bikin klaster baru nih #indonesiaterserah

Sebelum ke tahap *tokenization*, dilakukan pembersihan untuk membuat *string punctuation*, RT, simbol *hashtag* (#), simbol @, dan sebagainya. Kemudian, dilakukan tahap *tokenization*. *Tokenization* adalah mengubah kalimat menjadi per kata. Yang hasilnya dapat dilihat pada tabel 4.

Tabel 4. Hasil proses *tokenization*

Teks
['indonesia', 'hebat', 'kaya', 'dan', 'besar', 'hanya', 'saja', 'kebanyakan', 'manusianya', 'lemah', 'miskin', 'dan', 'kecil', 'dalam', 'berperilaku', '']
[' ', 'ini', 'bukan', 'soal', 'cebong', 'kampret', 'ini', 'soal', 'kemanusiaan', 'ini', 'soal', 'keadilan', 'ini', 'soal', 'pengorbanan', '']
['bakalan', 'bikin', 'klaster', 'baru', 'nih']

Proses selanjutnya adalah proses penghapusan *stop word*, yakni penghilangan kata yang tidak bermakna. Hasil dari proses ini dapat dilihat pada tabel 5. Kemudian dilanjutkan dengan proses *stemming*, yakni penghilangan imbuhan. Hasil dari proses ini ditunjukkan oleh tabel 6.

Dalam pemrosesan awal data, banyak data yang tersaring ketika dilakukan proses pembuangan duplikasi data sehingga pada tahap ini tersisa 236 *tweet* yang akan dianalisis. Duplikasi data dapat terjadi ketika seseorang melakukan *retweet* atau ketika proses *crawling* menggunakan API Twitter *free*.

Tabel 5. Hasil proses *stop word*

Teks
['indonesia', 'hebat', 'kaya', 'kebanyakan', 'manusianya', 'lemah', 'miskin', 'berprilaku', '']
['', 'cebong', 'kampret', 'kemanusiaan', 'keadilan', 'pengorbanan', '']
['bikin', 'klaster', 'nih']

Tabel 6. Hasil proses *stemming*

Teks
indonesia hebat kaya banyak manusia lemah miskin berprilaku
cebong kampret manusia adil korban
bikin klaster nih

Pembobotan Kata

Algoritma TF-IDF digunakan dalam pembobotan kata. Melalui algoritma ini diketahui kata yang paling banyak muncul di dokumen tertentu. Hasil dari proses pembobotan kata dapat dilihat melalui gambar 3. Kata-kata yang sering muncul adalah 'ya' dan 'hukum'.

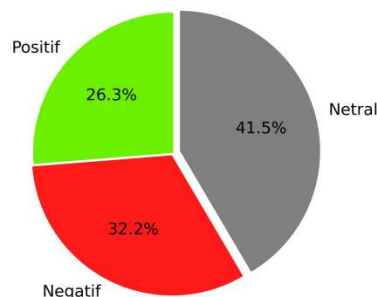
	term	rank
6	indonesia	4.510402
7	orang	4.547709
4	aja	4.728526
5	hukum	4.779013
8	ya	5.286022

Gambar 3. Hasil Pembobotan Kata

Pemodelan

Algoritma *sentistrength* digunakan dalam klasifikasi sentimen dari tagar #IndonesiaTerserah [11]. Dalam pengklasifikasian menggunakan algoritma *sentistrength*, akan dilakukan pencocokan hasil pemrosesan awal dengan isi leksikon. Jika kata hasil pemrosesan awal ada di leksikon maka akan ditampilkan bobot nilai dari kata tersebut. Hasil dari analisis ini akan menentukan berapakah polaritas negatif, positif, atau netral dari data. Polaritas tersebut didasarkan apabila lebih banyak *tweet* yang bernilai positif dibandingkan negatif maka akan menghasilkan polaritas bersentimen positif, begitu pula sebaliknya, apabila antara *tweet* bernilai positif dan negatif seimbang, maka akan menghasilkan sentimen bernilai netral.

Hasil Analisis Sentimen Dari 236 ungkapan pengguna



Gambar 4. Hasil Analisis Sentimen

Dari gambar 4 diketahui bahwa sebanyak 41,5% adalah netral, 32,2% negatif, dan 26,3% positif. Dari data tersebut dapat diketahui bahwa masyarakat banyak memberikan sentimental netral. Sentimen netral yang dihasilkan penelitian ini perlu dikaji lebih lanjut karena tidak bisa diyakini bahwa masyarakat Indonesia tidak tertarik dengan masalah penanganan virus Corona di Indonesia ataupun sudah tidak peduli dengan kebijakan yang diambil oleh pemerintah. Sesuai dengan yang disampaikan oleh dr. Tirta bahwa arti tagar ini bukan berarti tenaga kesehatan yang menyerah dalam menghadapi pandemi, tetapi mengarah pada

[illegible]

		Data Prediksi		
		Negatif	Netral	Positif
Data Aktual	Negatif	53	15	27
	Netral	20	81	17
	Positif	3	2	18

J ~ I C O N
urnal Komputer & Informatika

- [5] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, 2017, doi: [10.22146/ijccs.24716](https://doi.org/10.22146/ijccs.24716).
- [6] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 1, p. 21, 2020, doi: [10.24014/ijaidm.v3i1.9145](https://doi.org/10.24014/ijaidm.v3i1.9145).
- [7] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," 2016, doi: [10.21108/indosc.2016.111](https://doi.org/10.21108/indosc.2016.111).
- [8] M. N. Saadah, R. W. Atmagi, D. S. Rahayu, and A. Z. Arifin, "SISTEM TEMU KEMBALI DOKUMEN TEKS DENGAN PEMBOBOTAN TF-IDF DAN LCS," *JUTI J. Ilm. Teknol. Inf.*, 2013, doi: [10.12962/j24068535.v11i1.a16](https://doi.org/10.12962/j24068535.v11i1.a16).
- [9] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment in short strength detection informal text," *J. Am. Soc. Inf. Sci. Technol.*, 2010, doi: 10.1002/asi.21416.
- [10] "Confusion Matrix" . [Online]. Available: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> . [Accessed: 28-Des-2020].
- [11] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, 2016, doi: [10.22146/ijccs.16625](https://doi.org/10.22146/ijccs.16625)
- [12] "Jelaskan Soal Tagar #IndonesiaTerserah, dt. Tirta: Disini Bukan Kami Menyerah". [Online]. Available: <https://www.pikiran-rakyat.com/nasional/pr-01977403/jelaskan-soal-tagar-indonesia-terserah-dr-tirta-disini-bukan-kami-menyerah> . [Accessed: 28-Des-2020].