

Terbit : 30-September-2022

# Penerapan Metode Lexicon Based untuk Menganalisis Sentimen Terhadap Mudik Lebaran

Muhammad Wisnu Prayuda<sup>1)</sup>, Angga Aditya Permana<sup>2)\*</sup>

<sup>1)</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang, Banten, Indonesia

<sup>2)\*</sup>Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara, Banten, Indonesia

<sup>1)</sup>[Wisnu.prayuda@ft-umt.ac.id](mailto:Wisnu.prayuda@ft-umt.ac.id), <sup>2)\*</sup> [angga.permana@umn.ac.id](mailto:angga.permana@umn.ac.id)

## Abstrak :

Twitter adalah salah media sosial dengan pengguna terbanyak dari seluruh dunia dan di media sosial tersebut semua orang dapat mengungkapkan pendapat mereka. Pada tahun 2022, merupakan tahun yang dimana kasus penyakit yang sedang melanda seluruh bagian negara tidak terkecuali di Indonesia yaitu *Covid19* menurun dan hanya sedikit yang masih terjangkit penyakit tersebut. Dan di tahun tersebut merupakan pertama kalinya pemerintah Indonesia mengizinkan warganya untuk melakukan perjalanan jarak jauh dan kembali ke kampung halamannya masing-masing atau yang biasa disebut mudik. Penelitian ini diharapkan dapat bermanfaat membantu riset atas opini masyarakat yang mengandung sentimen positif, netral, dan negatif. Tahapan penelitian dilakukan dengan cara pengumpulan data *tweet* atau *crawling data*, preprosesing, lalu klasifikasi. Algoritma yang digunakan dalam penelitian ini menggunakan metode *Lexicon Based* dengan tingkat akurasi sebesar 85% dan menggunakan *tools* atau aplikasi *Python Notebook Jupyter*.

## Kata kunci :

komponen; Sentimen Analisis; Covid-19; Vaksin; Mudik; Lexion Based

## PENDAHULUAN

*World Health Organization* (WHO) mengumumkan lebih dari 52 juta orang dinyatakan positif Covid-19, dan 1.2 juta meninggal pada minggu kedua November 2020. Sementara itu, Indonesia mencatat 463 ribu orang dengan 15.148 kematian terkonfirmasi positif (Prabowo and Wiguna 2021). Media Sosial saat ini telah menjadi alat komunikasi yang sangat populer dari berbagai kalangan pengguna internet. Jutaan pesan muncul setiap harinya di situs *web* populer yang menyediakan layanan untuk perpesanan seperti *Twitter*, *Facebook*, dan *Instagram* (Syakuro 2017)(Budiman, Young, and Suryadibrata 2021). Pengguna Internet itu membagikan tentang kehidupan mereka, berbagi pendapat dengan berbagai topik dan mendiskusikan masalah yang ada saat ini. Karena format pesan yang gratis dan aksesibilitas yang mudah dari sosial media, pengguna internet saat ini cenderung beralih dari alat komunikasi tradisional ke layanan *microblogging* atau media sosial. Analisis Sentimen yaitu studi yang terdiri dari *Natural Language Processing*, komputasi linguistik, dan analisis teks untuk mengidentifikasi sentimen teks dapat membantu untuk mengidentifikasi opini mengenai suatu produk yang disampaikan oleh pengguna online untuk mengekspresikan emosi, perilaku, atau secara otomatis(Fauzi and Adinugroho 2018)(Destitus, Wella, and Suryasari 2020). Analisis sentimen biasanya terbagi menjadi tiga kelas yaitu positif, netral, dan negatif.

Analisis sentimen merupakan cabang dari *text mining*, fokus utamanya adalah menganalisa dokumen teks. Dokumen teks terkadang mengandung unsur non-tekstual, salah satunya emoji. Emoji merupakan simbol grafis *Unicode* berupa gambar untuk mengekspresikan perasaan seseorang. Algoritma yang digunakan dalam penelitian ini adalah *Lexicon Based* dengan pembaharuan berupa penambahan pembobotan non-tekstual (emoji). Hasil dari pembobotan tekstual dan non tekstual yang dinormalisasi dengan metode *Confusion Matrix* digabungkan disertai nilai konstanta tertentu sehingga menghasilkan sentimen positif, netral maupun negatif(Lestari, Perdana, and Fauzi 2017).

\*Angga Aditya Permana

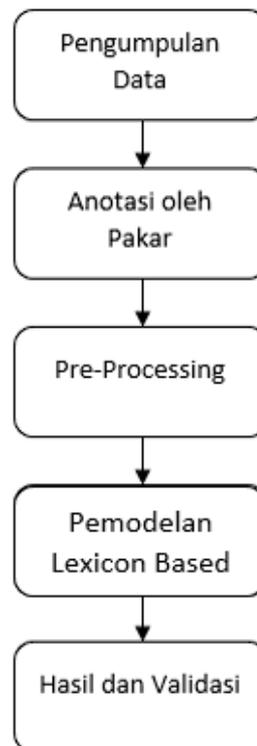


Metode *Lexicon-Based* merupakan salah satu metode ekstraksi fitur yang dapat meningkatkan performansi fitur (Azhar 2018). Metode ini bekerja dengan cara membuat kamus berupa kata terlebih dahulu, lalu kata-kata tersebut digunakan untuk mengidentifikasi apakah kata-kata tersebut mengandung opini atau tidak (Rosadi et al. 2021).

Salah satu media sosial yang dimana banyak pengguna dari penjuru dunia dapat mengungkapkan perasaan ataupun pendapat mereka yaitu *Twitter* (Zuhdi, Utami, and Raharjo 2019). Data yang dihasilkan di *Twitter* juga bisa sangat berguna jika dianalisa karena data tersebut dapat diekstrak menjadi informasi penting melalui *opini mining*. Pendapat tentang berita atau peluncuran produk apapun atau bahkan jenis tren tertentu dapat diamati dengan baik di *Twitter*.

### METODE PENELITIAN

Metode penelitian yang dilakukan pada penelitian ini memiliki 5 tahapan: Pengumpulan Data, Anotasi Pakar, *Pre-Processing*, Pemodelan Klasifikasi Lexicon Based serta Hasil dan Validasi seperti pada gambar di bawah ini.



Gambar 1. Metodologi Penelitian

Proses ini dimulai dengan sebuah proses *Pengumpulan* data *tweet* dengan cara *crawling*, lalu proses *anotasi* dari seorang annotator kemudian di lanjutkan dengan proses *pre-processing* yang berfungsi untuk membersihkan dataset lalu dilakukan *pemodelan klasifikasi NBC* dan di akhiri oleh *validasi*.

#### Pengumpulan Data

Dalam pengumpulan data, peneliti menggunakan media sosial *twitter* untuk mengambil Dokumen berupa *tweet* atau komentar dari akun *Twitter* yang berupa opini. Data *tweet* tersebut didapat dengan memanfaatkan fitur *API (Application Programming Interface)* yang telah disediakan oleh *twitter* itu sendiri. Dokumen yang dimasukkan merupakan dokumen berbahasa Indonesia dengan menggunakan metode *crawling data*.

\*Angga Aditya Permana



### Anotasi

untuk menciptakan hasil klasifikasi yang valid, maka dalam proses anotasi, dilakukan kerjasama penelitian dengan seorang ahli linguistik Bahasa Indonesia dan Bahasa Inggris untuk mengenali pesan yang berpotensi sebagai cyberbullying berdasarkan unsur perbuatan yang dilanggar berdasarkan pendekatan ilmu linguistik. Anotasi bertujuan untuk membangun knowledge pada model classifier (Nugraha Manoppo and Hatta Fudholi 2021) dan juga digunakan untuk validasi hasil klasifikasi. Anotasi membutuhkan annotator, yakni orang yang berkapasitas untuk melakukan anotasi sesuai dengan keperluan penelitian. Kapasitas annotator menjadi salah satu penentu model penelitian menjadi valid dan tidak mengalami kesalahan metodologi.

### Pre-Processing

Salah satu hal penting dalam penelitian ini adalah tahap preprocessing untuk ke tahap selanjutnya, untuk mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi. Data yang dimasukkan pada tahap ini masih berupa data mentah yang belum tersaring (*cleansing*) sehingga menghasilkan dokumen yang berkualitas dengan harapan mempermudah dalam proses klasifikasi.

### Pemodelan Lexicon Based

Berdasarkan uraian sebelumnya, data *tweet* didapat dengan menggunakan teknik *crawling* atau *scrapping data*. Penulis mendapatkan data sebanyak 389 *tweets* yang tersimpan dalam *database* yang berupa file *spreadsheet csv*. Proses pengambilan data menggunakan *API Twitter* yang diambil melalui *website developer* milik *twitter* itu sendiri. Data secara otomatis akan dikumpulkan secara otomatis berdasarkan *hashtag* dan tanggal pengambilan data yaitu dengan *hashtag* 'mudik' dengan pengambilan data pada periode 26 April 2022 hingga sekarang.

### Hasil dan Validasi

Teknik validasi dirancang untuk mendeteksi kesalahan umum yang mungkin terjadi. Biasanya hasil validasi didesain untuk menunjukkan alasan mengapa suatu nilai data ditandai. Ketika menentukan prosedur validasi yang akan diterapkan pada suatu variabel tertentu, keakuratan variabel mana yang dapat diamati dan mampu untuk mengoreksi kesalahan yang terdeteksi harus selalu diterapkan. (Windatiningsih 2019)

## HASIL PENELITIAN DAN DISKUSI

Tabel 4. 1 Rumus Confusion Matriks

		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

Tabel 4. 2 Data Sementara Setelah Scrapping

Username	Tweetcreatedts	text
ngaanuuu	2022-07-23 23:29:04	@inaco_jelly @SmgMenfess2 Home rasa away sisan mudik ke mertua ~
adeanita40	2022-07-23 23:20:01	RT @PSambadha: Waktu itu kita packing barang2 pribadi beliau caranya ya gak jauh beda dg kita semua. Barang2 kita packing rapi dalam dus2 bā€
heyalfen	2022-07-23 23:12:42	mudik <a href="https://t.co/agjAKWgQfP">https://t.co/agjAKWgQfP</a>

\*Angga Aditya Permana



Kolom *username* menunjukkan nama pengguna *twitter*, kolom *tweetcreatedts* menunjukkan waktu *tweet* tersebut di *posting* ke media sosial, dan text berisikan data *tweet* yang di ketik oleh pengguna media sosial *Twitter*. Selanjutnya hasil pengumpulan data *tweet* tersebut dijadikan dataset dan digunakan dalam proses analisa sentimen.

### 1. Hasil Preprocessing

Setelah tahap pengumpulan data, tahap selanjutnya adalah *preprocessing*. Berdasarkan uraian sebelumnya tahap ini terdiri dari *cleaning*, *case folding*, *stemming*, *tokenizing*, dan *stopword removal*. Hasil *preprocessing* secara garis besar data tweet sudah tidak mengandung kata-kata yang tidak perlu dalam melakukan sentimen seperti menghilangkan ‘RT’, menghilangkan username, menghilangkan simbol-simbol, menghilangkan *URL link*, dan lain-lain.

Tabel 4. 3 Data Preprocessing

Data Sebelum <i>Preprocessing</i>	Data Sesudah <i>Preprocessing</i>
Mantan gua jaman kuliah, tengah malem Tlpn cerita kesedihan RTnya, Suami Mandul, Nafkah harian, suruh mijit smp subuh lah, gua hiburan, gua tawarin tiket kalo dia mau refreshing kerumah Ortu nya. Giliran mudik ke Jakarta ga bilang2 ama gua, anjing ga niat Ngebadut malah dibadutin..	mantan jaman kuliah tengah malem tlpn cerita sedih rtnya suami mandul nafkah hari suruh mijit smp subuh lah hiburan tawarin tiket mau refreshing rumah ortu gilir mudik jakarta bilang ama anjing niat ngebadut malah dibadutin
w pernah punya mantan yang literally dia tanya ""mobilmu apa"" waktu itu kek liburan puasa hari raya gitu deh terus mudik kan yaa chat biasa gitu diperjalanan, terus dia tanya gitu kek anjir apaaa iniiii $\delta\dot{Y}\sim\delta\dot{Y}\sim\delta\dot{Y}\sim$ <a href="https://t.co/k8jvixGYz">https://t.co/k8jvixGYz</a>	w pernah punya mantan literally tanya mobil apa waktu libur puasa hari raya gitu deh terus mudik yaa chat biasa gitu jalan terus tanya gitu anjir apaaa iniiii
@RedDoorzID #ReminBagiGopay pingin mudik ke mertua di kabupaten cianjur sekalian ke kebun raya bogor sekalian biar hemat diongkos dan doakan menang kak @RedDoorzID	reminbagigopay pingin mudik mertua kabupaten cianjur sekali kebun raya bogor sekali biar hemat ongkos doa menang kak

Tabel 4. 4 Hasil dan Label Score

Hasil <i>Preprocessing</i>	Hasil Translate	Sentimen Score	Label sentimen
lockey indonesia ribu join kv fest pantau an nambah gak masuk list lah konser perdana tour perd	lockey indonesia thousands joined kv fest, monitoring and adding it was not included in the list, the first concert tour perd	0.0	Netral
mantan jaman kuliah tengah malem tlpn cerita sedih rtnya suami mandul nafkah hari suruh mijit smp subuh lah hiburan tawarin tiket mau refreshing rumah ortu gilir mudik jakarta bilang ama	ex-college in the middle of the night, a sad story calls, rt, husband is barren, he makes a living, he tells me to go to school at dawn, entertain him, offer a ticket, want to refresh his parents' house, take turns going	-0,128	Negatif

\*Angga Aditya Permana



anjing niat ngebadut malah dabadutin	home in Jakarta, saying that the dog intends to be a clown instead		
ayah temenku mau mudik kota c buat jenguk istri sama anak temen aku ayah pulang apotek aku coba kontak temenku buat nanya ayah tinggal dom temenku jawab iya ayah mas kerja sini terus aku tanya	my friend's father wants to go back to city c to visit my friend's wife and children my father comes home from the pharmacy I try to contact my friend to ask my father stay at my friend's office my friend answered yes my father works here and I ask	0.8442	Positif
bagus lha g perlu kena rasisme mau hal thailand ngonten skalian mudik skalian kambek skalian konser	It's good, you don't need to be subject to racism, positive thinking, want things from Thailand, you guys, you're content, you're going home, you're overcrowded, you're going to concerts	0.4019	Positif

Sebelum masuk ke tahap selanjutnya, data dari hasil tahap *Preprocessing* dapat menghasilkan kata yang sering muncul. Data tersebut dijadikan tumpuan relevansi untuk penulis antara kata yang sering muncul dengan *hashtag* atau kata yang di gunakan pada saat pengumpulan data yaitu kata 'mudik'. Pada kata mudik dalam data tersebut menunjukkan kata 'mudik' muncul sebanyak 396 kata.

## 2. Hasil Analisa Sentimen

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 adalah 0,55, score 0,55 termasuk dalam label positif karena batas ambang nilai score > 0. Contoh score adalah 0,0, score 0,0 termasuk dalam label netral karena batas ambang nilai score = 0. Contoh score adalah -0,55, score -0,55 termasuk dalam label negatif karena batas ambang nilai score < 0.

Hasil klasifikasi menggunakan TextBlob menghasilkan kelas sentimen positif lebih banyak dari hasil pelabelan manual, jumlah kelas sentiment positif dari klasifikasi TextBlob berjumlah 389.

Tabel 4. 5 Perhitungan Confusion Matrix

		Data Prediksi		
		Positif	Netral	Negatif
Data Aktual	Positif	181	117	12
	Netral	253	156	3
	Negatif	19	121	52

$$Akurasi = \frac{52+156+181}{500} = 0,778 = 77.8\%$$

Tabel 4. 6 Hasil Presisi, Recal, F1-Score

Kelas	Presisi	Recall	F1-Score	Akurasi
Positif	0.88	0.84	0.86	0.86
Netral	0.85	0.85	0.85	
Negatif	0.80	0.92	0.86	

\*Angga Aditya Permana





- 11(2):107–11. doi: 10.31937/si.v11i2.1740.
- Fauzi, Muhammad Ali, and Sigit Adinugroho. 2018. "Analisis Sentimen Pariwisata Di Kota Malang Menggunakan Metode Naive Bayes Dan Seleksi Fitur Query Expansion Ranking Optimasi Sisa Bahan Baku Pada Industri Mebel Menggunakan Algoritma Genetika View Project Automatic Essay Scoring View Project." (August). Lestari, Agnes Rossi Trisna, Rizal Setya Perdana, and M. Ali Fauzi. 2017. "Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dan Pembobotan Emoji." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 1(12):1718–24.
- Nugraha Manoppo, Tommy, and Dhomas Hatta Fudholi. 2021. "Deteksi Cyberbullying Berdasarkan Unsur Perbuatan Pidana Yang Dilanggar Dengan Naive Bayes Dan Support Vector Machine." *Jurnal Sains Komputer & Informatika (J-SAKTI)* 5(1):10.
- Prabowo, Wahyu Adi, and Citra Wiguna. 2021. "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM." *Jurnal Media Informatika Budidarma* 5(1):149. doi: 10.30865/mib.v5i1.2604.
- Rosadi, Aqwam, Dede Gustiana, Manajemen Informatika, Stmik Jakarta Sti, Jl B. R. I. No, Radio Dalam, Keb Baru, Jakarta Selatan, Perbaikan Kata, Tidak Baku, and Mengubah Kata Bernegasi. 2021. "Analisis Sentimen Berdasarkan Opini Pengguna Pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based Dan Naive Bayes Classi Er Twitter Text Mining." 20:39–52.
- Syakuro, Abdan. 2017. "Pada Media Sosial Menggunakan Metode Naive Bayes Classifier ( NBC ) Dengan Seleksi Fitur Information Gain ( IG ) Halaman Judul Skripsi Oleh : Abdan Syakuro." *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)* 1–89.
- Windatiningsih, Desi. 2019. "Uji Validasi Data Debit." *Jurnal Sumber Daya Air* 15(2):121–36. doi: 10.32679/jsda.v15i2.600.
- Zuhdi, Abdul Malik, Ema Utami, and Suwanto Raharjo. 2019. "Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN." *Jurnal Informa Politeknik Indonusa Surakarta* 5:1–7.

\*Angga Aditya Permana



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.