

Analisis Sentimen Opini Publik di platform X (Twitter) terhadap Bencana Banjir Sumatra

Andi Laksono¹, Ferdhian Akbar Rahmadani²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *1andilaksana125@gmail.com, *2ferdhianakbar19@gmail.com

Abstrak – Indonesia merupakan wilayah yang rentan terhadap bencana hidrometeorologi, salah satunya banjir yang secara berulang menimbulkan dampak sosial dan ekonomi yang signifikan. Pada Desember 2025, banjir besar yang melanda berbagai wilayah di Pulau Sumatra memicu peningkatan aktivitas komunikasi publik di media sosial, khususnya platform X (Twitter), sebagai sarana penyampaian informasi, keluhan, dan opini secara real-time. Kondisi ini menjadikan media sosial sebagai sumber data potensial untuk memahami dinamika respons dan persepsi masyarakat selama krisis bencana. Namun, karakteristik data Twitter yang bersifat tidak terstruktur, berukuran besar, serta sarat dengan variasi bahasa informal menimbulkan tantangan dalam proses analisis. Penelitian ini memanfaatkan pendekatan Natural Language Processing untuk menganalisis sentimen opini publik terhadap bencana banjir di Sumatra dengan menggunakan representasi fitur Bag of Words yang diperkuat pembobotan Inverse Document Frequency. Data diperoleh melalui proses harvesting tweet menggunakan Tweet Harvest dan dianalisis menggunakan perangkat lunak Orange Data Mining dengan alur kerja yang sistematis dan transparan. Pendekatan ini memungkinkan pemetaan kecenderungan sentimen publik serta eksplorasi kata-kata kunci yang dominan muncul dalam percakapan daring terkait bencana, sehingga memberikan gambaran awal mengenai respons psikologis masyarakat pada situasi krisis kebencanaan.

Kata Kunci – analisis sentimen, bencana, opini publik, orange data mining, twitter

1. PENDAHULUAN

Indonesia secara geografis terletak pada wilayah yang sangat rentan terhadap bencana hidrometeorologi, di mana banjir menjadi fenomena tahunan yang memberikan dampak masif pada sektor ekonomi dan sosial. Pada pengujung tahun 2025, intensitas curah hujan yang ekstrem telah memicu banjir besar di berbagai provinsi di Pulau Sumatra, yang mengakibatkan kerusakan infrastruktur, memaksa ribuan warga untuk mengungsi bahkan menimbulkan kematian. Dalam situasi krisis seperti ini, pola komunikasi masyarakat mengalami pergeseran ke arah digital, di mana platform media sosial X (dahulu Twitter) menjadi salah satu kanal utama bagi publik untuk menyampaikan informasi terkini, keluhan, maupun aspirasi secara real-time [1][2]. Pemanfaatan data dari media sosial ini memberikan peluang besar bagi otoritas terkait untuk memahami dinamika psikologi massa dan kebutuhan mendesak di lapangan melalui analisis opini yang tersebar secara luas.

Namun, volume data yang sangat besar dan sifat teks yang tidak terstruktur pada platform X menjadi tantangan tersendiri dalam proses ekstraksi informasi. Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing* (NLP) yang mampu mengklasifikasikan opini masyarakat ke dalam kategori tertentu untuk melihat kecenderungan respon publik terhadap penanganan bencana [3]. Salah satu komponen krusial dalam NLP adalah pembobotan kata untuk menentukan kepentingan relatif sebuah kata dalam dokumen. Metode *Inverse Document Frequency* (IDF) digunakan dalam penelitian ini untuk memfilter kata-kata yang umum muncul dan menonjolkan istilah-istilah unik yang memiliki nilai informasi tinggi terkait bencana banjir Sumatra, sehingga hasil klasifikasi menjadi lebih akurat dan representatif [4][5].

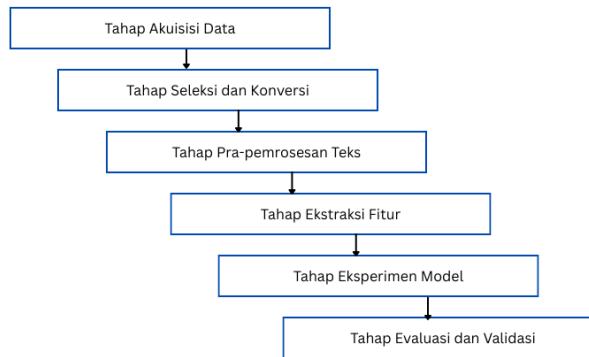
Penelitian terdahulu menunjukkan bahwa penggunaan perangkat lunak berbasis visual seperti *Orange Data Mining* sangat efektif dalam memproses data mining yang kompleks dengan alur kerja yang transparan dan *reproducible* [6]. Meskipun banyak penelitian analisis sentimen telah dilakukan, namun studi yang berfokus pada peristiwa spesifik seperti bencana dengan memanfaatkan instrumen *Tweet-Harvest* masih sangat terbatas. Penggunaan *Tweet-Harvest* memungkinkan perolehan dataset yang lebih komprehensif tanpa batasan yang sering ditemui pada API resmi platform X (twitter) yang sangat terbatas, sehingga menyediakan data primer yang lebih kaya untuk dianalisis [7].

Penelitian ini bertujuan untuk memetakan sentimen opini publik terhadap bencana banjir di Sumatra pada Bulan Desember 2025 kemarin dengan memanfaatkan *Orange Data Mining* sebagai alat analisis utama. Selain menerapkan pendekatan *Bag of Words* (IDF) sebagai representasi pembobotan fitur kata yang dominan, penelitian ini juga melakukan komparasi kinerja berbagai algoritma klasifikasi guna mendapatkan model dengan akurasi

terbaik. Urgensi dalam membandingkan berbagai metode ini sejalan dengan studi komparasi algoritma [8], yang menegaskan pentingnya evaluasi performa model seperti *Random Forest*, *SVM*, *Decision Tree* dan sebagainya untuk memastikan ketepatan prediksi. Dengan demikian, hasil analisis ini diharapkan tidak hanya memberikan kontribusi teoritis pada bidang informatika khususnya *Natural Language Processing* (NLP) dan analisis sentimen, tetapi juga memberikan kontribusi praktis sebagai bahan evaluasi bagi pemerintah dan lembaga kemanusiaan dalam merumuskan strategi komunikasi krisis yang lebih responsif dan tepat sasaran.

2. METODE PENELITIAN

Kerangka kerja penelitian ini disusun berdasarkan alur kerja sistematis yang mengintegrasikan proses akuisisi data mentah hingga evaluasi model klasifikasi. Alur ini direpresentasikan secara visual melalui alur kerja pada aplikasi *Orange Data Mining* yang terdiri dari tahapan sebagai berikut:



Gambar 1. Alur Kerja Penelitian

2.1 Tahap Akuisisi Data (*Data Harvesting*)

Proses dimulai dengan pengambilan data primer menggunakan tool *Tweet-harvest* berbasis *Node.js*. Pengumpulan data difokuskan pada platform X dengan parameter kata kunci "bencana" yang terjadi di wilayah Sumatra atau Indonesia selama periode Tahun 2025.

2.2 Tahap Seleksi dan Konversi (*Selection & Corpus*)

Data mentah dalam format CSV dimuat melalui *widget File*. Implementasi dilanjutkan dengan penggunaan *widget Select Columns* untuk menetapkan fitur teks dan *created at* sebagai atribut *metas* atau teks. *Widget Select Rows* digunakan untuk memfilter data yang hanya pada tahun 2025 karena tool *Tweet-harvest* terbaru tidak bisa melakukan filter tanggal dimulai dan tanggal berakhir pencarian, lalu kemudian dikonversi menjadi objek *Corpus* agar dapat diproses oleh algoritma *Text Mining*.

2.3 Tahap Pra-pemrosesan Teks (*Preprocessing*)

Pada tahap ini, dilakukan pembersihan data secara intensif melalui *Widget Preprocess Text*. Langkah-langkahnya meliputi *tokenization* (pemecahan kata), *case folding* (penyeragaman huruf kecil), dan penghapusan attribute *url* dan *hyperlink*. Selain itu, diterapkan *stopword removal* untuk mengeliminasi kata-kata umum yang tidak memiliki nilai informatif dalam konteks bencana. Juga ditambahkan filter lain seperti *Lexicon Default*, *Filter Number* dan penghapusan karakter non-alfabet atau *regexp*.

2.4 Tahap Ekstraksi Fitur (*Bag of Words & IDF*)

Proses ini didukung oleh analisis paralel melalui *widget Sentiment Analysis* untuk melihat *Sentiment Distribution*, *Word Cloud* dan *Topic Modelling* untuk melihat kata yang sering muncul, dan juga *Corpus* yang telah bersih ditransformasikan menjadi matriks numerik menggunakan model *Bag of Words* (BoW). Untuk mengoptimalkan relevansi kata kunci, diterapkan metode pembobotan *Inverse Document Frequency* (IDF). Secara teknis, IDF akan memberikan bobot tinggi pada istilah-istilah spesifik terkait bencana yang jarang muncul di dokumen lain namun memiliki signifikansi tinggi bagi klasifikasi. Lalu dilakukan *Tweet Profiling* dan *Select Emotion Column* untuk melihat Distribusi Emosi Pengguna Twitter (X) dan juga peta sebaran Menggunakan *Box Plot*.

2.5 Tahap Eksperimen Model (*Modelling*)

Implementasi pemodelan dilakukan dengan membandingkan lima algoritma klasifikasi secara simultan untuk menemukan performa terbaik. Algoritma yang digunakan meliputi *Random Forest*, *Naive Bayes*, *Decision Tree*,

SVM dan *Neural Network*. Yang mana data emosi berasal dari pra pemrosesan *Bag of Words* yang sudah dilakukan Filter menggunakan *Tweet Profiler* untuk Menentukan Emosi dari Pengguna.

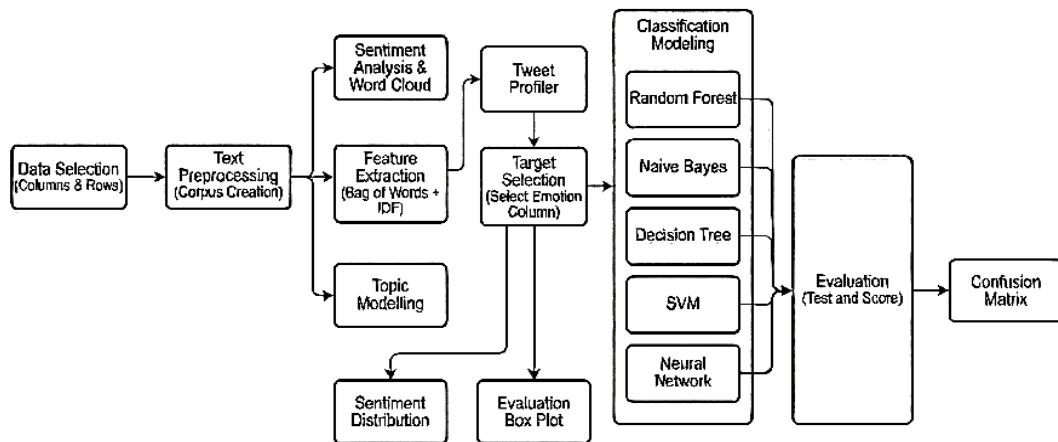
2.6 Tahap Evaluasi dan Validasi (*Evaluation*)

Alur kerja diakhiri dengan pengujian model menggunakan *widget Test and Score* dengan metode *Cross-Validation* sebanyak 5 kali. Hasil performa dari kelima algoritma divalidasi melalui *Confusion Matrix* untuk memetakan sebaran sentimen secara eksplisit.

3. HASIL DAN PEMBAHASAN

3.1 Skenarion Penelitian

Dalam penelitian ini digunakan aplikasi *orange data mining* versi 3.40. dimana desain penelitian dari *orange data mining* sebagai berikut:



Gambar 2. Desain Penelitian

A. *Crawling/Harvest* Data X (Twitter)

Proses pengumpulan data menggunakan *tools Tweet-Harvest* guna alternatif pembatasan API resmi *twitter* yang membatasi hanya sebanyak 100 *tweet* perbulan. Proses pengumpulan data pada media sosial X (*twitter*) dilakukan dengan metode *web scraping* melalui chromium dengan fitur auto scrolling kata pencarian yang sudah diberikan.

```
ACER JAWA ID@ACERJAWAID-HAHAHAHHAH MINGW64 ~/Desktop/tweet-harvest-main
$ npx tweet-harvest@latest
Tweet Harvest [v2.7.1]
Tweet Harvest [v2.7.1]

Research by Helmi Satria
Use it for Educational Purposes only!

This script uses Chromium Browser to crawl data from Twitter with your Twitter
auth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.
Note: This script only runs on your local device.

✓ What's your Twitter auth token? ... *****
✓ What's the search keyword? ... bencana
✓ How many tweets do you want to crawl? ... 800
✓ What format do you want to export? » CSV

Opening twitter search page...
```

Gambar 3. Proses *Crawler* Data Twitter

Dalam proses tersebut diberikan beberapa ketentuan seperti banyaknya data tweet yang dicari. Tetapi menurut penelitian yang sudah dilakukan, per hari data tweet yang bisa dipanen hanya sejumlah kurang dari 700 data tweet, dikarenakan kebijakan terbaru dari platform media sosial tersebut. Dan setelah data tweet didapatkan sesuai kata kunci yang diberikan data disimpan dalam bentuk CSV guna dilakukan analisis dalam aplikasi *Orange Data Mining*.

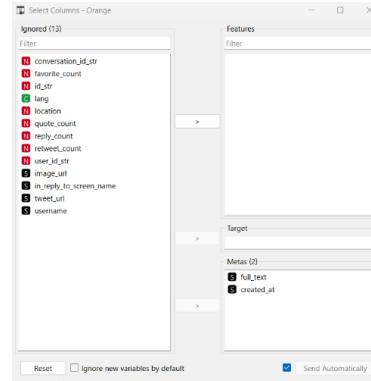
crested_at	full_text	
494 - 5-12-2018	Libur man, ketiga puj oneng. Cluban asalain momen berlangsung bersama keluarga, pastikan kesempatan rehat jadi prioritas.Diputuh zulmu di lokasi wisata, lengkap diri dengan perlengkapan kesehatan, serta ...	
495 - 5-12-2018	Membuka buku keran perang, mungkin akhirnya kerana Bangzai (Dianah) ini korah jadi salah satu yang 1.896 orang. https://www.kutafit.id/Ualgz	
496 - 5-12-2018	Dilengah sifat kawau-kawau di Sumatra merangkap kerajaan, takdirnya tak fikir masing dan tuntut esok.	
497 - 5-12-2018	Has #makanDikasih#Dikemas#Pemerasan yang merupakan pengaruh bantuan logistik dan kebutuhan dasar ke seluruh tempat terkena. Pengeluar hilangklik dilakukan untuk mempercepat distribusi bantuan.	
498 - 5-12-2018	Siap-siap! Gerak cepat pemerintah untuk korban banjir Sumatra Selatan di kabupaten tanpa adanya fasilitas yang memadai agar korban bantuan bisa diterima.	
499 - 5-12-2018	Pemerintah menginginkan semua Sumberdaya Bantuan dapat berfungsi dengan baik melalui penyaluran bencana di Aceh, Sumbar & Sumut, saingi kita baka serba, saling menerima, saling jag... 500 - 5-12-2018	Hentikan kebodohan "hukum eksterior". Sumberdaya Bantuan dapat berfungsi dengan baik melalui penyaluran bencana di Aceh, Sumbar & Sumut, saingi kita baka serba, saling menerima, saling jag... 501 - 5-12-2018
502 - 5-12-2018	Iku stadia jadi bencana nacionals! Logistik lebih cepat penanganannya! (Dilanjut melihat... https://www.kutafit.id/pejepet)	
503 - 5-12-2018	Mendengar Tito Kurniawan ketekunan tentang rasa nuansa kotak bencana yang lengkapnya. Hal ini disampaikan Tito di Kabupaten Tapamau-Tengah, Sumatra Utara (Sumut). Menggi (2/12/2018) DHA... 504 - 5-12-2018	MEMAHAMI "KEGERAMAN" MENANTU UNHURU BANJARANTAU (Dilanjut melihat... https://www.kutafit.id/pejepet)
505 - 5-12-2018	"Banyak yang tahu tentang banjir, tetapi tidak pernah merasakannya." Poin Munirah Putri dalam wawancara dengan dilansir oleh https://www.kutafit.id/pejepet	
506 - 5-12-2018	506 - 5-12-2018 - Agi... (Agi) hujan di akhir ramen ni. Dijamin ilang atau, yang ke arah si apal, injinca dan rindine everyday, gitali werning atau pun, parau, bencana di "mari" dan para kondo kondo ni ikumannya memaki, lo... 507 - 5-12-2018	Hal ini dalam artian dalam Pasal 281 UU, yang merupakan hal-hal teknis seperti crach bantah untuk memperoleh dan menyampaikan informasi. Detil konten bantah, pembenaran informasi dan sebagainya.
508 - 5-12-2018	Bencana merupakan faktor krisis, tak seperti mengakibatkan krisis. Atau memang orang yang memberi wacana tentang amalan seperti Sebutlah terdiri dari sejati seputar ikutan tentang... lapis tent... 509 - 5-12-2018	Sebenarnya bagian ngeleti story teman-teman yang turun langsung dari relasi bencana di Sumatra... Senangnya saat nanti bisa jadi relawan bencana juga, Aamiin!
510 - 5-12-2018	Aksi Warga Jadi Perpanjangan Pera Bencana" (Liputan6.com) Warga Tulus Wajibkan Benar Alami Mengjerap Anginan dan Jalin Silaturahmi Di Desa Cengkong Kec. Sabungan kab. Tapanuli Selatan (Dinas Kesiwa dan Penerbangan)	
511 - 5-12-2018	Lapek, para jadi perpanjangan rasa ngidap, Dzidz. Ko laper je, tetapi rada sulit wasup merasakan kekayaan bumi Indonesia yg ditambahkan lagi dengan makanan yg.. DZIDZIING makanan yg bukan, dibatasi oleh, resutan makan... 512 - 5-12-2018	Nama nama orang penting yang akan terkena sanksi atau sanksi atas bantuan yang memerlukan sanksi adalah di Pulu Sumatra, KORPRI Bandar burung bergerak mengadakan dzidzidziing diDZIDZIING tersebut ultimatum... 513 - 5-12-2018
514 - 5-12-2018	Had sebabu blintut, utegge wadup kewapulan dan solatian, atas bencana yang memerlukan sanksi adalah di Pulu Sumatra, KORPRI Bandar burung bergerak mengadakan dzidzidziing diDZIDZIING tersebut ultimatum... 515 - 5-12-2018	Bencana berasal karena bencana bersebutan dan dideklarasikan oleh organisasi pelajar, keselamatkan raya wakaf dianggap, dzidzidziing makanan manusia Indonesia
516 - 5-12-2018	Pemerintah mengeluarkan surat edaran mengenai penanganan bencana di wilayah Sumatera Barat. Surat edaran ini bertujuan untuk memberikan petunjuk dan perintah mengenai penanganan bencana di wilayah Sumatera Barat.	
517 - 5-12-2018	Teman Widi jadi Pak Bojang di Tembaga Dinchia (BD) atau bengkong je di rumah je berbulut dan berbadan je. https://www.kutafit.id/pejepet	
518 - 5-12-2018	Dilengah kepentingan medan dan situasi kreatifitas, negara tetep hidup. Heliopatra TH Asiaquest dan Angkatan regim 101/12/2018 merelaksasikan diri silih sunyi demikian teman-teman he... 519 - 5-12-2018	Negara selalu berusaha menciptakan medan bagi faktor alam agar benci kepada dirinya. DZIDZIING bantuan kerja di desa setelah ketahuan oleh zon, akan alih bentuknya menjadi mana yang hasilnya bertanggung jawab...
520 - 5-12-2018	Sehu, bermula emas malah kambang obat-obatan bercita-cita dan perasaan ketakutan, menteri sifatnya like a hero yang hilang neigris. Alok sedar sekarang akar kiyati semestu setara	
521 - 5-12-2018	Kapit PT Pemda Pembuatan dan Logistik Pulp Berasih, Persepatan Bencana di Sumatra Utara https://www.kutafit.id/pejepet	
522 - 5-12-2018	Ada fenomena menakjubkan mutu berasih yang diakui. In ikut semesta tentang negara, pentingnya mengambil tentang kiat yang memiliki umat dzidzidziing kebaka metaku TNI-Polri, metaku pertamina yang L...	
523 - 5-12-2018	Nabu "malapetaka" atau "bencana besar", menjuluki peristiwa 1948, ketika 700.000 warga Palestina mengungsi diDzidzidziing 1948 terpaci karena negara? Arab memulai perang melawan Israel, hasilnya, kalah diDzidzidziing 20...	

Gambar 4. Contoh Data *Tweet* hasil *Crawling*

Terlihat bahwa data didapatkan sejumlah 523 data *tweet* dengan tanggal pencarian yang beragam dikarenakan keterbatasan dari tool yang digunakan.

B. Data Selection (Columns & Rows)

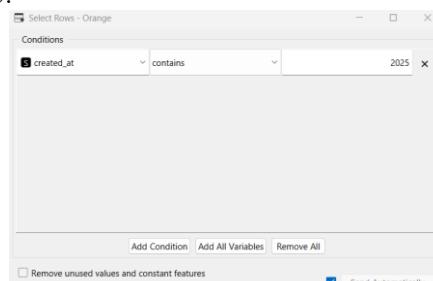
Tahap analisis dimulai dengan impor dataset hasil *crawling* dari *tools Tweet-Harvest*. Pada widget ini, dilakukan seleksi atribut (*feature selection*) untuk memilih kolom yang relevan bagi penelitian, seperti teks tweet dan tanggal tweet dibuat, serta membuang fitur yang tidak diperlukan (seperti *user_id*, *conversation_id*, *url*, *reply*, *quote* dan sebagainya). Proses ini memastikan bahwa hanya akan memproses data yang memiliki signifikansi terhadap konteks bencana khususnya Banjir di Sumatra.



Gambar 5. Seleksi Kolom

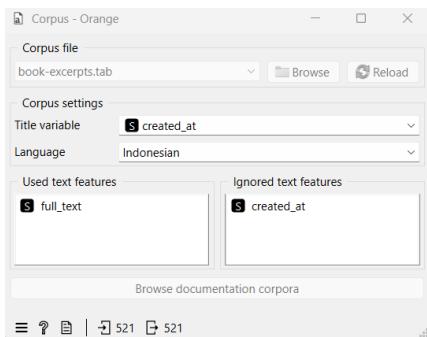
Gambar 3. Seleksi Kelim

Setelah dilakukan seleksi pada kolom dilakukan pula seleksi pada baris dengan hanya menggunakan kata kunci 2025 sebagai filter agar tweet yang diproses hanya yang bertahun 2025 saja. Dan menghasilkan sejumlah 521 data tweet yang dibuat di tahun 2025.



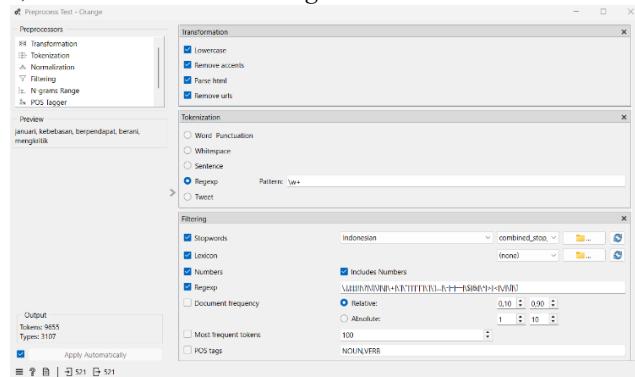
the unused classes

C. Text Preprocessing (Corpus Creation)
Data teks mentah dari Twitter yang bersifat tidak terstruktur ditransformasikan menjadi format yang siap



Gambar 7. Pemilihan Kolom Corpus

Setelahnya dilakukan proses *preprocessing* meliputi serangkaian teknik pembersihan data dalam standar NLP, antara lain: *Transformation*, *Tokenization* dan *Filtering*

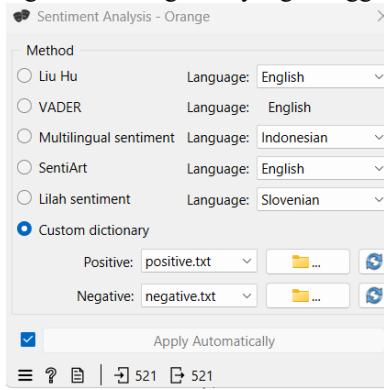


Gambar 8. Pemilihan Fitur dalam Preprocessing Text

Tahapan ini sangat krusial untuk mereduksi noise pada data. Hasil akhirnya adalah sekumpulan token kata bersih yang merepresentasikan inti percakapan mengenai bencana, yang kemudian menjadi basis bagi analisa dan ekstraksi fitur selanjutnya.

D. Sentiment Analysis, Word Cloud dan Topic Modelling

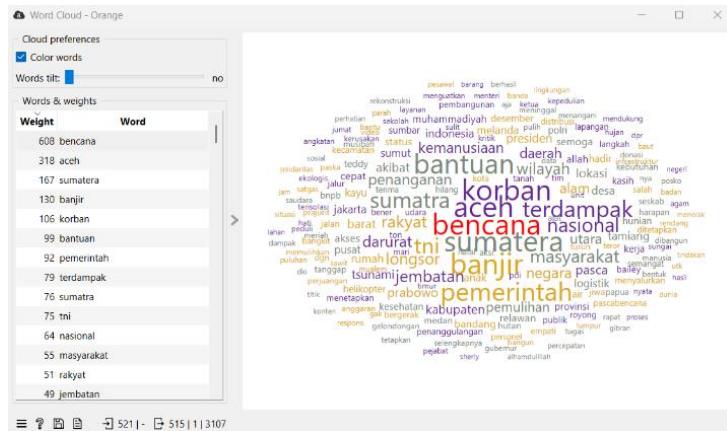
Secara paralel, analisis sentimen awal dilakukan untuk memberikan label polaritas pada setiap dokumen (positif atau negatif). Langkah ini memberikan gambaran umum mengenai persepsi psikologis masyarakat terhadap penanganan banjir sebelum masuk ke pemodelan yang lebih kompleks. Disini peneliti menggunakan *custom dictionary* yang terdapat di situs web github tentang NLP yang menggunakan Bahasa Indonesia.



Gambar 9. Pemilihan Fitur dalam Preprocessing Text

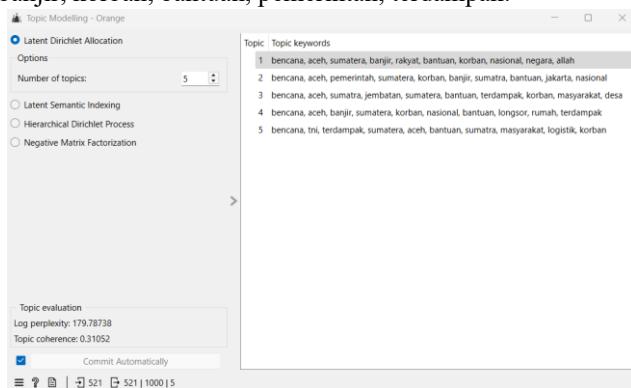
Lalu didapatkan data sentimen negatif sejumlah 358, data sentimen netral sejumlah 87 dan data sentimen positif sejumlah 76. Atau jika diuraikan dalam bentuk persen, sentimen negatif sejumlah 68,71%, netral sejumlah 16,70% dan positif sejumlah 14,59%.

Sebagai bagian dari analisis data, dilakukan visualisasi *Word Cloud* untuk memetakan kata-kata yang paling dominan muncul dalam diskusi publik. Hal ini membantu peneliti untuk segera mengidentifikasi topik utama (misalnya: "bencana", "banjir", "korban", "bantuan").



Gambar 10. Analisa dengan *Word Cloud*

Topic Modelling (seperti *Latent Dirichlet Allocation*) digunakan secara terpisah untuk mengelompokkan tweet ke dalam tema-tema laten yang tersembunyi. Terlihat dari *Word Cloud* kata yang paling sering muncul adalah bencana, aceh, sumatra, banjir, korban, bantuan, pemerintah, terdampak.



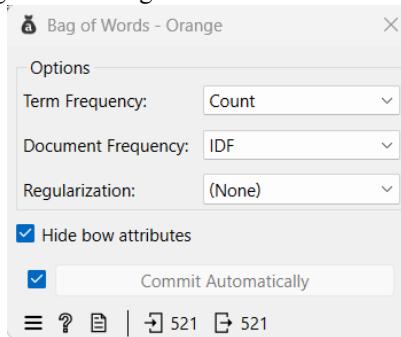
Gambar 10. Analisa dengan *Topic Modelling*

Dalam *Topic Modelling* yang dibatasi sejumlah lima topik Terlihat ada Kata yang paling sering muncul antara lain :

1. bencana, aceh, sumatera, banjir, rakyat, bantuan, korban, nasional, negara, allah
 2. bencana, aceh, pemerintah, sumatera, korban, banjir, sumatra, bantuan, jakarta, nasional
 3. bencana, aceh, sumatra, jembatan, sumatera, bantuan, terdampak, korban, masyarakat, desa
 4. bencana, aceh, banjir, sumatera, korban, nasional, bantuan, longsor, rumah, terdampak
 5. bencana, tni, terdampak, sumatera, aceh, bantuan, sumatra, masyarakat, logistik, korban

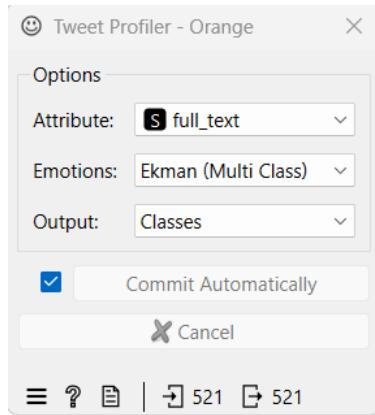
E. Feature Extraction (*Bag of Words*:IDF) dan Tweet Profiler

Transformasi teks menjadi data numerik dilakukan menggunakan metode *Bag of Words*. Pada penelitian ini, pembobotan fitur diperkuat dengan penerapan *Inverse Document Frequency* (IDF). Opsi IDF dipilih untuk memberikan bobot lebih tinggi pada kata-kata unik yang spesifik muncul dalam konteks bencana, dan memberikan penalti (bobot rendah) pada kata-kata yang terlalu sering muncul secara umum.



Gambar 11. Setting Parameter BoW

Tweet Profiler mengambil data opini pada setiap dokumen dan akan menghitung probabilitas atau nilai sentimen. Widget ini mendukung 3 kategori yaitu *Ekman*, *Plutchiks* dan *Mood State Profiles* (POMS). Dalam penelitian ini digunakan klasifikasi *Ekman* dengan opsi *Multi class* sebagai berikut.:

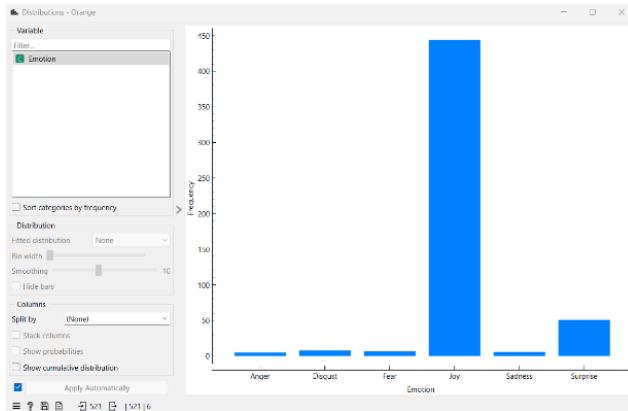


Gambar 12. Tweet Profiler

Widget *Tweet Profiler* digunakan untuk memperkaya data dengan mendeteksi emosi spesifik (seperti takut, marah, sedih, senang) yang terkandung dalam teks guna melengkapi analisa sentiment lanjutan.

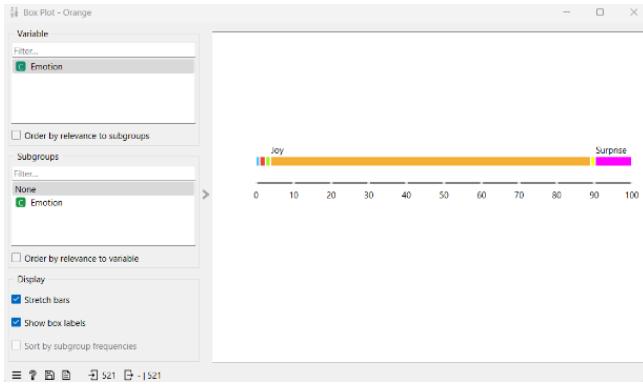
F. Sentiment Distribution & Evaluation Box Plot

Sebelum pelatihan model, distribusi data target dianalisis menggunakan *Sentiment Distribution* dan *Box Plot*. Berdasarkan grafik *Sentiment Distribution* yang dihasilkan, terlihat adanya fenomena menarik di mana kategori emosi 'Joy' (Sukacita) mendominasi secara signifikan dibandingkan emosi negatif seperti 'Fear' (Takut) atau 'Anger' (Marah). Secara kuantitatif, frekuensi 'Joy' melampaui 400 instance, sementara kategori lain berada di bawah angka 50.



Gambar 13. Tabel Distribusi

Visualisasi *Box Plot* mengonfirmasi temuan pada distribusi frekuensi, di mana sebaran data terkonsentrasi kuat pada kategori 'Joy' (diveisualisasikan dengan batang oranye panjang) dan sebagian kecil pada 'Surprise' (Kejutan).

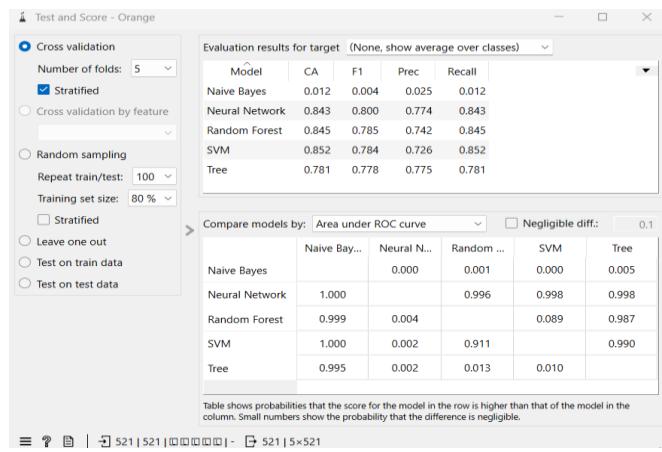


Gambar 14. Tabel Box Plot

G. Evaluation (Test and Score)

Evaluasi kinerja model dilakukan menggunakan metode validasi silang (*k-fold cross-validation*), umumnya dengan *default* aplikasi $k=5$. Metode ini memastikan bahwa model diuji pada subset data yang berbeda-beda untuk

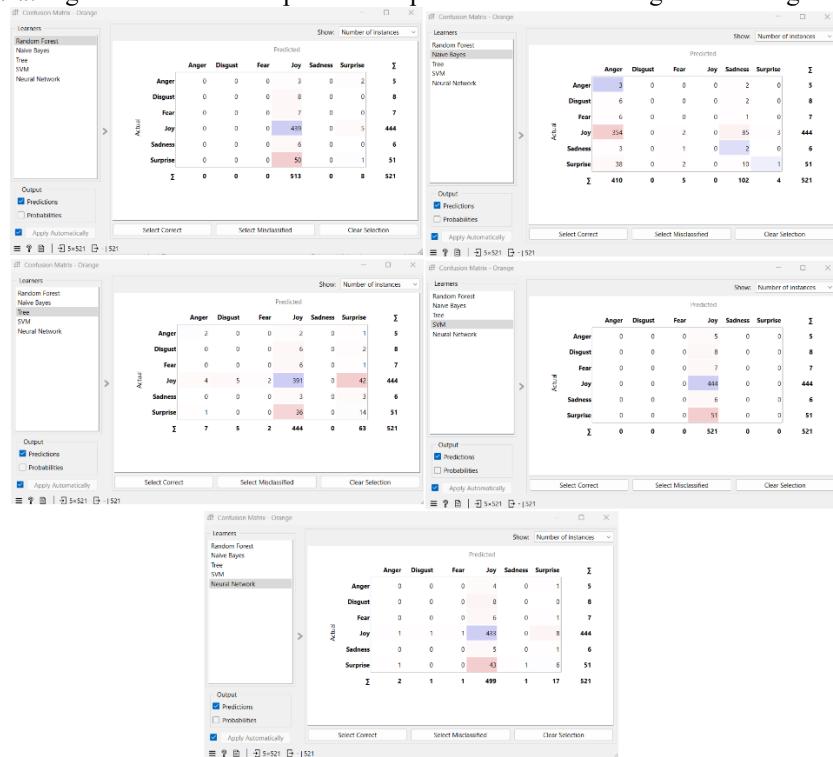
menghindari *overfitting*. Metrik evaluasi yang diukur meliputi *Classification Accuracy (CA)*, *F1-Score*, *Precision*, dan *Recall*. Hasil dari Test dan Score adalah sebagai berikut :



Gambar 15 Hasil Test and Score

H. Confusion Matrix

Confusion Matrix digunakan untuk memperlihatkan prediksi dari kelima algoritma sebagai berikut:



Gambar 16 Confusion Matriks

Dan hasil dari *Confusion Matriks* Menunjukan Hasil Sebagai Berikut :

Tabel 1 Hasil *Confusion Matriks*

Emosi	Actual		Random Forest		Naïve Bayes		Decision Tree		SVM		Neural Network	
Anger (Kemarahan)	5	1%	0	0%	410	79%	7	1%	0	0%	2	0%
Disgust (Jijik)	8	2%	0	0%	0	0%	5	1%	0	0%	1	0%
Fear (Ketakutan)	7	1%	0	0%	5	1%	2	0%	0	0%	1	0%
Joy (Kegembiraan)	444	85%	513	98%	0	0%	444	85%	521	100%	499	96%
Sadness (Kesedihan)	6	1%	0	0%	102	20%	0	0%	0	0%	1	0%
Surprise (Kejutan)	51	10%	8	2%	4	1%	63	12%	0	0%	17	3%

3.2 Pembahasan

Berdasarkan hasil evaluasi menggunakan metode *Cross Validation* pada *tabel Test and Score*, algoritma *Support Vector Machine (SVM)* mencatatkan performa akurasi (*Classification Accuracy/CA*) tertinggi sebesar 0.852, disusul sangat ketat oleh *Random Forest* (0.845) dan *Neural Network* (0.843). Tingginya akurasi pada ketiga model ini (di atas 84%) menunjukkan bahwa model mampu memprediksi kelas mayoritas (emosi ‘Joy’) dengan sangat baik. Secara khusus, SVM terbukti efektif dalam menangani data teks berdimensi tinggi hasil pembobotan BoW (IDF), di mana ia mampu memisahkan *hyperplane* antar kelas sentimen secara optimal. Namun, jika meninjau metrik *F1-Score* yang merepresentasikan keseimbangan antara *Precision* dan *Recall*, *Neural Network* justru unggul dengan skor 0.800, lebih tinggi dibandingkan SVM (0.784). Hal ini mengindikasikan bahwa meskipun SVM lebih akurat secara global, *Neural Network* memiliki kemampuan generalisasi yang sedikit lebih baik dalam mengenali kelas-kelas minoritas.

Di sisi lain, terdapat anomali signifikan pada kinerja algoritma *Naive Bayes* yang mencatatkan akurasi sangat rendah, yakni 0.012 atau 1.2%. Kegagalan masif pada model ini kemungkinan besar disebabkan oleh ketidakmampuan algoritma *Naive Bayes* dalam menangani asumsi independensi fitur pada data teks yang sangat tidak seimbang (*highly imbalanced dataset*) seperti yang terlihat pada distribusi emosi sebelumnya. Ketimpangan drastis antara performa *Naive Bayes* dengan model berbasis *ensemble* (*Random Forest*) atau *network* (*Neural Network*) menegaskan bahwa untuk dataset bencana banjir Sumatera ini, pendekatan non-linear dan kompleksitas model sangat diperlukan untuk menangkap pola sentimen yang sarat nuansa. Oleh karena itu, model seperti SVM atau *Neural Network* direkomendasikan sebagai model utama untuk dipublikasikan dan dipergunakan.

4. SIMPULAN

Dominasi label emosi ‘Joy’ pada visualisasi *Sentiment Distribution* sesungguhnya menyingkap paradoks ironis dalam respons publik terhadap krisis kebencanaan. Alih-alih memicu gelombang kritik dan kemarahan publik (*Anger*) atau menuntut pertanggungjawaban struktural atas bencana yang terjadi, data justru menunjukkan sikap naif masyarakat yang cenderung permisif dan merayakan “keselamatan parsial”. Tingginya sentimen ini mengindikasikan bahwa publik terjebak dalam pola pikir *toxic positivity*, di mana fokus attensi bukan tertuju pada akar masalah kegagalan mitigasi, melainkan teralihkan pada euforia sesaat atas bantuan logistik atau narasi heroisme relawan yang sebenarnya tidak menyelesaikan masalah mendasar.

Fenomena ini merefleksikan rendahnya ekspektasi publik dan nalar kritis masyarakat terhadap standar pelayanan negara dalam penanggulangan bencana. Emosi sukacita yang muncul merupakan bentuk romantisasi penderitaan, di mana masyarakat secara tidak sadar menormalisasi situasi darurat sebagai “takdir” yang harus diterima dengan lapang dada (*nrimo ing pandum*), bukan sebagai kegagalan tata kelola infrastruktur. Akibatnya, data ini menjadi sinyal peringatan bahwa tekanan sosial (*social pressure*) yang diperlukan untuk mendorong perbaikan kebijakan publik justru tumpul karena tertutup oleh rasa syukur sesaat yang tidak ditempatkan pada konteks yang semestinya.

5. SARAN

Penelitian selanjutnya disarankan untuk tidak hanya berhenti pada klasifikasi emosi umum (seperti *Joy* atau *Sadness*), melainkan menerapkan *Aspect-Based Sentiment Analysis* (ABSA). Hal ini penting untuk membedah lebih dalam ke arah mana emosi tersebut ditujukan; apakah ‘Joy’ tersebut ditujukan pada keselamatan diri (*relief*), atau kinerja pemerintah (*satisfaction*). Dengan ABSA, peneliti dapat memisahkan antara rasa syukur personal dengan kepuasan terhadap layanan publik, sehingga interpretasi mengenai ‘sikap naif’ masyarakat dapat diverifikasi lebih akurat.

Mengingat tingginya kemungkinan penggunaan gaya bahasa satir atau dark jokes oleh netizen Indonesia dalam merespons bencana, penelitian mendatang perlu mengintegrasikan modul deteksi sarkasme (*Sarcasm Detection*). Selain itu, pengayaan kamus (*lexicon*) yang mengakomodasi istilah lokal atau slang spesifik (seperti ‘nrimo’, ‘pasrah’, ‘gws negara’) sangat diperlukan agar algoritma tidak salah mengklasifikasikan ungkapan keputusasaan atau sindiran halus sebagai emosi positif rasa senang (*Joy*).

Disarankan untuk melakukan analisis time-series dengan rentang waktu yang lebih panjang, mencakup fase pra-bencana, tanggap darurat, hingga pasca-pemulihan. Penelitian ini hanya memotret momen krisis (antara bulan Desember tahun 2025), sehingga ada kemungkinan emosi ‘Joy’ hanyalah euforia sesaat (*short-term relief*). Studi lanjutan perlu melihat apakah sentimen positif tersebut bertahan atau justru berubah menjadi kemarahan (*Anger*) ketika bantuan logistik berhenti dan masalah struktural mulai terasa kembali.

DAFTAR PUSTAKA

- [1] F. Fathoni, A. Ibrahim, F. R. Mumtaz, M. A. Zaky, M. J. Pratama, and I. A. Kurniawan, “ANALISIS SENTIMEN PUBLIC TWITTER TERHADAP KEBIJAKAN PEMERINTAH MENGGUNAKAN METODE SVM (STUDI KASUS : RUU TNI),” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 4, pp. 6322–6329, May 2025, doi: 10.36040/JATI.V9I4.14036.
- [2] A. M. Maksun, Y. A. Sari, and B. Rahayudi, “Analisis Sentimen pada Twitter Bencana Alam di Kalimantan Selatan menggunakan Metode Naive Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5614–5621, Nov. 2021, Accessed: Dec. 31, 2025. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10307>
- [3] G. Prima Ertansyah, R. Tri, C. Kusuma, and A. A. Sari, “Analisis Sentimen Pada Media Sosial Menggunakan Teknik Natural Language Processing,” *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 183–189, Jul. 2025, doi: 10.47701/QGCEY104.
- [4] F. S. Widyastuti and E. Mailoa, “ANALISIS SENTIMEN ULASAN KONSUMEN MENGGUNAKAN ALGORITMA TF-ID UNTUK MENGETAHUI TINGKAT KEPUASAN PELANGGAN(STUDI KASUS : GUNTHEM PREMIUM COFFEE),” *J. Teknol. Inf. DAN Komun.*, vol. 16, no. 2, pp. 8–16, Sep. 2025, doi: 10.51903/JTIKP.V16I2.1010.
- [5] K. T. Putra, M. A. Hariyadi, and C. Crysdiyan, “Perbandingan Feature extraction TF-IDF dan BOW Untuk Analisis Sentimen Berbasis SVM,” *J. Cahaya Mandalika ISSN 2721-4796*, vol. 3, no. 2, pp. 1449–1463, 2022, Accessed: Dec. 31, 2025. [Online]. Available: <https://www.ojs.cahayamandalika.com/index.php/jcm/article/view/2292>
- [6] Ismail, H. N. Rahmah, and R. Sulistiyowati, “PENGUNAAN SOFTWARE ORANGE DATA MINING PADA IMPLEMENTASI TEXT MINING DALAM ANALISIS SENTIMEN NETIZEN DI TWITTER TERHADAP KELANGKAAN MINYAK GORENG,” *Sigma-Mu*, vol. 14, no. 2, pp. 1–11, Sep. 2022, doi: 10.35313/SIGMAMU.V14I2.4667.
- [7] R. Ramadhani, T. Hidayat, and S. Sukisno, “Klasifikasi ANALISIS SENTIMEN TWEET MASYARAKAT PADA ISU INDONESIA GELAP DI MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN METODE COUNTVECTORIZER,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 6, pp. 9664–9670, Nov. 2025, doi: 10.36040/JATI.V9I6.15729.
- [8] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, “Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia,” *Semin. Nas. Teknol. Sains*, vol. 2, no. 1, pp. 377–382, Jan. 2023, doi: 10.29407/STAINS.V2I1.2877.