

Analisis Sentimen Potensi Gempa *Megathrust* di Indonesia Menggunakan *Random Forest*

Mikhael Christian^{1*}, Agung Panji Sasmito¹, Hani Zulfia Zahro¹

¹Program Studi S1 Teknik Informatika, Institut Teknologi Nasional Malang, Indonesia

*Email: 2118006@scholar.itn.ac.id

Info Artikel	Abstrak
<p>Kata Kunci : analisis sentimen, <i>random forest</i>, media sosial x, <i>megathrust</i></p> <p>Keywords : <i>sentiment analysis, random forest, x social media, megathrust</i></p> <p>Tanggal Artikel Dikirim : 25 November 2024 Direvisi : 16 Desember 2024 Diterima : 10 Februari 2025</p>	<p>Pada bulan Agustus tepatnya di tahun 2024, berita mengenai adanya potensi gempa <i>megathrust</i> di Indonesia membuat kalangan warga internet heboh salah satunya di media sosial X. Hal ini menimbulkan kekhawatiran dan adanya perbedaan sudut pandang dalam menanggapi berita ini di kalangan masyarakat terlebih para pengguna media sosial. Dikarenakan Indonesia sendiri memiliki 13 zona subduksi yang tersebar di beberapa wilayah. Zona subduksi sendiri merupakan pertemuan antar lempeng besar yang saling bertubrukan dan dapat mengakibatkan gempa <i>megathrust</i>. Akan tetapi, terdapat dua zona yang hangat diperbincangkan, yaitu <i>Megathrust Mentawai-Siberut</i> dan <i>Selat Sunda</i>. Penelitian ini bertujuan untuk mengevaluasi reaksi masyarakat <i>online</i> dalam menanggapi isu <i>megathrust</i>. Pendekatan yang akan digunakan dalam penelitian ini adalah metode <i>Random Forest</i>, yang dikenal mampu menangani data yang besar dan akurasi yang dihasilkan tinggi. Tahapan pengolahan data meliputi pembersihan teks (<i>cleaning</i>), <i>case folding</i>, normalisasi, tokenisasi, penghapusan <i>stopword</i>, <i>stemming</i>, serta adanya evaluasi model. Data yang digunakan untuk melakukan pelatihan dan pengujian berasal dari <i>tweet</i> dengan kata kunci <i>megathrust</i>. Hasil pengujian menggunakan data latih sebanyak 136 data dan menghasilkan nilai <i>recall</i> 84.38%, <i>precision</i> sebesar 71.05%, dan <i>accuracy</i> sebesar 63.24% dalam mengklasifikasikan sentimen netral. Penelitian ini diharapkan dapat membantu lembaga pemerintahan dan organisasi sadar bencana memahami persepsi masyarakat di dunia maya. Dengan wawasan ini, lembaga terkait dapat lebih efektif dalam merancang strategi komunikasi, penyuluhan, serta mitigasi risiko, sehingga mendukung peningkatan kesiapsiagaan bencana di Indonesia.</p>
	<p>Abstract</p> <p><i>In August, in 2024, news of a potential megathrust earthquake in Indonesia caused a stir among internet citizens, including on social media X. This caused concern and different perspectives on this news among the public, especially social media users. This caused concern and different points of view in responding to this news among the public, especially social media users. Indonesia itself has 13 subduction zones spread across several regions. The subduction zone itself is a meeting between large plates that collide with each other and can cause megathrust earthquakes. However, there are two zones that are hotly discussed, namely the Mentawai-Siberut Megathrust and the Sunda Strait. This study aims to evaluate the reaction of online communities in responding to the megathrust issue. The approach that will be used in this research is the Random Forest method, which is known to be able to handle large data and high accuracy. Data processing stages include text cleaning, case folding, normalization, tokenization, stopword removal, stemming, and model evaluation. The data used to conduct training and testing comes from tweets with megathrust keywords. The test results used 136 training data and produced a recall value of 84.38%, precision of 71.05%, and accuracy of 63.24% in classifying neutral sentiment. This research is expected to help government agencies and disaster awareness organizations understand public perception in cyberspace. With this insight, related institutions can be more effective in designing communication strategies, counseling, and risk mitigation, thus supporting the improvement of disaster preparedness in Indonesia.</i></p>

1. PENDAHULUAN

Pada bulan Agustus tahun 2024, pengguna media sosial X digemparkan dengan adanya berita mengenai potensi gempa *megathrust* di Indonesia. Gempa *megathrust* adalah jenis gempa bumi yang terjadi akibat pergerakan lempeng berukuran luas yang saling bertubrukan yang disebut sebagai zona subduksi. *Megathrust* dapat terjadi karena mayoritas wilayah Indonesia berada di lokasi pertemuan tiga lempeng tektonik, yaitu Lempeng Indo-Australia di sebelah barat, Lempeng Eurasia di utara, dan Lempeng Pasifik di timur [1]. Menurut informasi Pusat Studi Gempa Nasional, Indonesia dikelilingi oleh 13 zona subduksi yang diperkirakan memiliki kekuatan lebih dari M8.

Namun, BMKG memperingatkan bahwa kita harus mewaspadai gempa dari dua segmen subduksi berikut, yaitu Megathrust Mentawai-Siberut dan Megathrust Selat Sunda. Hal ini dikarenakan pada kedua zona ini telah mengalami *seismic gap* selama 227 tahun dan 267 tahun [2], sedangkan biasanya gempa besar memiliki siklus tersendiri dalam rentang hingga ratusan tahun [3]. *Seismic Gap* merupakan area patahan tektonik yang aktif, akan tetapi telah lama atau jarang mengalami gempa, sehingga bisa berpotensi mengakibatkan gempa besar [4]. Berdasarkan penelitian [5] kita perlu mengawasi *seismic gap* mengingat gempa yang dihasilkan berskala besar.

Hingga saat ini, tidak sedikit dari warga internet yang merasa khawatir dan panik usai beredarnya informasi mengenai potensi gempa megathrust yang dapat mengancam beberapa wilayah di Indonesia [6]. Di sisi lain, terdapat juga masyarakat yang menganggap bahwa peringatan dini mengenai gempa megathrust sangatlah penting [7]. Walaupun begitu, BMKG mengimbau baik masyarakat ataupun warga internet untuk tetap tenang dan melanjutkan aktivitas sehari-hari seperti biasanya [8]. Namun, beragamnya respons masyarakat ini menunjukkan adanya kebutuhan untuk memahami sentimen secara komprehensif. Sayangnya, sebagian besar penelitian sebelumnya cenderung fokus pada analisis sentimen secara umum tanpa mempertimbangkan isu kebencanaan yang memerlukan pendekatan khusus, terutama untuk memahami persepsi masyarakat dalam konteks mitigasi risiko.

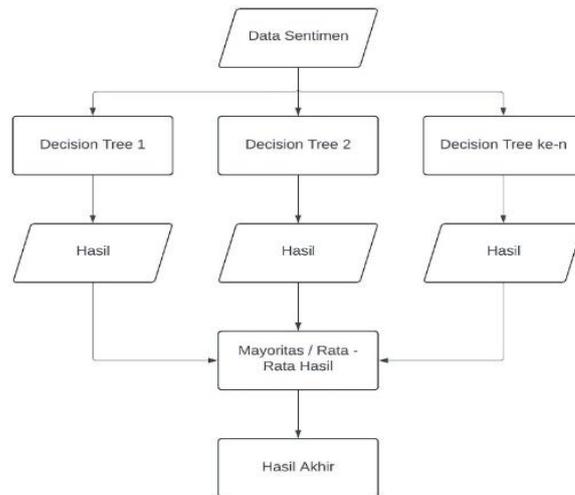
Penelitian ini mengisi gap tersebut dengan memanfaatkan analisis sentimen berbasis data untuk mengklasifikasikan sudut pandang masyarakat terhadap isu gempa *megathrust*. Proses analisis sentimen pada penelitian ini menggunakan metode *Random Forest*. Pilihan metode ini didorong oleh kemampuannya dalam memproses data dengan volume besar serta memberikan hasil akurasi yang tinggi [9]. Pada penelitian [10], dikatakan bahwa algoritma *Random Forest* merupakan suatu bentuk lanjutan dari algoritma *Decision Tree*. Metode *Random Forest* juga mendapatkan hasil akurasi yang tinggi sebesar 93% saat dilakukan klasifikasi analisis sentimen terhadap review pengguna aplikasi JMO [11]. Akan tetapi, terdapat kelemahan pada metode *Random Forest* yaitu penciptaan banyaknya tree yang membutuhkan daya komputasi lebih besar serta berpengaruh dalam lamanya proses pelatihan data [12].

Sehubungan dengan itu, tujuan penelitian ini adalah untuk melaksanakan analisis dan mengklasifikasikan sudut pandang warga internet baik secara positif, negatif, maupun netral dalam beropini mengenai berita potensi gempa *megathrust* di Indonesia. Melalui penelitian yang dilakukan, diharapkan dapat meningkatkan kesadaran bagi lembaga pemerintahan maupun komunitas sadar bencana tentang persepsi dan respons masyarakat terutama warga internet terkait isu tersebut, sehingga dapat digunakan untuk melakukan penyuluhan, mitigasi risiko, dan pengambilan keputusan yang lebih tepat.

2. METODE PENELITIAN

2.1 Metode *Random Forest*

Metode klasifikasi *Random Forest* merupakan algoritma klasifikasi yang menggunakan teknik *ensemble learning*, dimana algoritma tersebut menggabungkan beberapa model dari algoritma *Decision Tree* yang bertujuan untuk menghasilkan hasil prediksi yang lebih baik. Pada penelitian ini juga dikatakan bahwa algoritma *Random Forest* adalah model lanjutan dari algoritma *Decision Tree*, sehingga apabila ada penambahan *dataset*, maka *tree* yang terdapat di dalam metode *Random Forest* juga akan ikut berkembang [13]. Algoritma *Random Forest* dapat dilihat pada Gambar 1.



Gambar 1. Algoritma *Random Forest*

Algoritma *Random Forest* merupakan algoritma *machine learning* yang ditujukan untuk menjalankan regresi maupun klasifikasi. Algoritma ini mengembangkan sejumlah pohon keputusan yang dibangun berdasarkan kumpulan acak dari data *training* [14].

2.2 TF-IDF

Term Frequency-Inverse Document Frequency atau TF-IDF adalah salah satu prosedur yang digunakan untuk menentukan nilai bobot pada suatu kata [15]. Perhitungan nilai *Term Frequency* dan *Inverse Document Frequency* akan dilakukan untuk setiap dokumen dalam korpus yang ada, atau dengan kata lain, untuk menentukan seberapa sering dan pentingnya suatu kata dalam dokumen tersebut [16]. Berikut adalah persamaan algoritma TF-IDF pada persamaan (1).

$$W = tf \times idf \quad (1)$$

Keterangan:

W = Bobot TF-IDF

TF = Hasil kalkulasi nilai *Term Frequency*

IDF = Hasil kalkulasi nilai *Inverse Document Frequency*

2.3 Term Frequency (TF)

Perhitungan *Term Frequency* (TF) diterapkan untuk mengetahui kemunculan suatu kata dalam sebuah dokumen [17]. Semakin besar perhitungan yang dihasilkan, maka frekuensi kemunculan kata tersebut dalam dokumen semakin banyak. Persamaan *Term Frequency* (TF) dapat dilihat pada persamaan (2).

$$Tf(w) = \frac{\text{Frekuensi muncul kata } w \text{ di dokumen } d}{\text{Total kata pada dokumen } d} \quad (2)$$

2.4 Inverse Document Frequency (IDF)

Penerapan perhitungan *Inverse Document Frequency* (IDF) dilakukan untuk mengukur kepentingan sebuah kata. Berbanding terbalik dengan TF, jika sebuah kata jarang muncul maka nilainya justru semakin besar [18]. Hal ini juga ditujukan untuk mengurangi bobot suatu kata jika kemunculannya terlalu banyak. Persamaan *Inverse Document Frequency* (IDF) dapat dilihat pada persamaan (3).

$$IDF(w) = \log_e \frac{\text{Jumlah total dokumen}}{\text{Total dokumen pada kata } w} \quad (3)$$

2.5 Evaluasi Model

Tahap ini diadakan untuk mengukur keefektifan algoritma *Random Forest* dalam proses pengklasifikasian sentimen pengguna media sosial. *Confusion Matrix* merupakan salah satu sarana yang digunakan agar peneliti dapat menilai seberapa efektif sebuah model klasifikasi dalam mengidentifikasi entri dari berbagai kelas [19]. Dengan menggunakan *confusion matrix* [20] akan dilakukan beberapa perhitungan seperti pengukuran akurasi yang dapat dilihat pada persamaan (4).

$$\text{Accuracy} = \frac{(TP + TN)}{\text{Total Data}} \times 100\% \quad (4)$$

Keterangan:

- a. *True Positives* (TP)
Merujuk ke sentimen yang diprediksi sebagai positif dan dinyatakan benar positif.
- b. *True Negatives* (TN)
Merujuk ke sentimen yang diprediksi sebagai negatif dan dinyatakan benar negatif.
- c. *False Positives* (FP)
Merujuk ke sentimen yang diprediksi sebagai positif dan dinyatakan negatif.
- d. *False Negatives* (FN)
Merujuk ke sentimen yang diprediksi sebagai negatif dan dinyatakan positif.

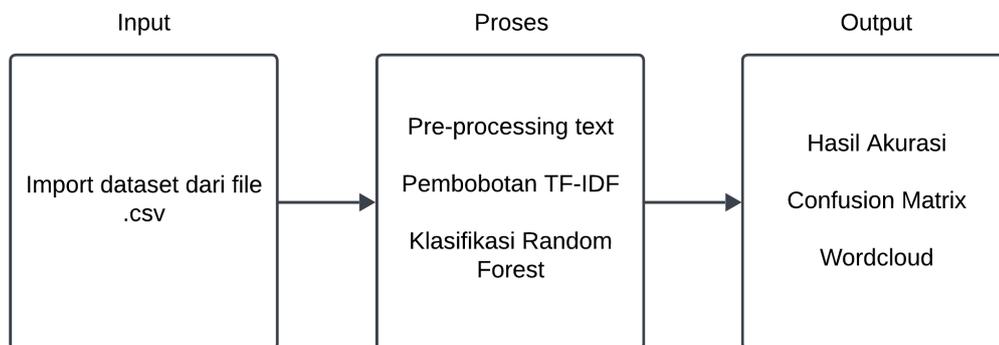
Kemudian akan dilakukan perhitungan *Precision* untuk melihat akurasi antara data yang diminta dan hasil prediksi yang dihasilkan oleh model. Perhitungan ini dapat dilihat pada persamaan (5).

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\% \quad (5)$$

Terakhir adalah dilakukannya perhitungan *Recall* untuk menunjukkan seberapa baik model dalam mengenali dan mengklasifikasikan data ke dalam kelas positif dengan benar. Perhitungan ini dapat ditemukan pada persamaan (6).

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100\% \quad (6)$$

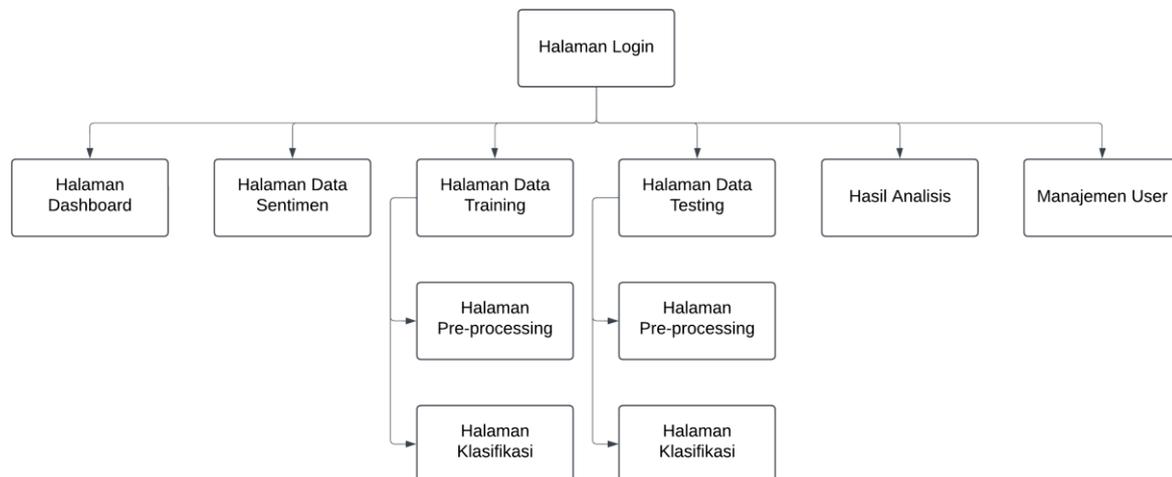
2.6 Diagram Blok Sistem



Gambar 2. Blok Diagram Sistem

Gambar 2. merupakan diagram blok yang mencakup proses inputasi yaitu *import dataset* dengan bentuk *file csv*. Kemudian proses selanjutnya yaitu *pre-processing text*, pembobotan TF-IDF, dan klasifikasi menggunakan model *Random Forest*. Setelah dilakukan proses tersebut, terdapat *output* pada sistem berupa hasil akurasi, *confusion matrix*, dan *wordcloud*.

2.7 Struktur Menu

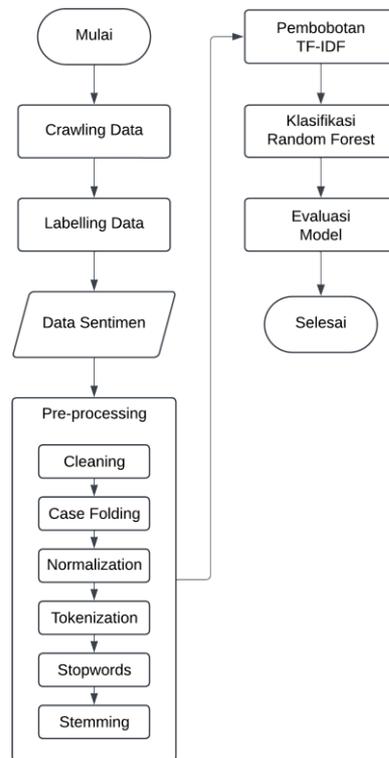


Gambar 3. *Flowchart* Struktur Menu

Gambar 3. di atas adalah struktur menu *super admin* yang dapat mengakses semua halaman. Pertama super admin akan masuk pada halaman *login*, kemudian terdapat beberapa halaman, yaitu halaman *dashboard*, halaman data sentimen, halaman data *training*, halaman data *testing*, halaman hasil analisis, dan halaman manajemen *user*. Kemudian pada halaman data *training* dan data *testing* dibagi menjadi dua, yaitu halaman *pre-processing* dan halaman klasifikasi. Pemilihan rasio data *training* 90% dan data *testing* 10% ditujukan agar model mampu mempelajari pola data yang terbatas secara menyeluruh. Pada penelitian [21] dikatakan bahwa pembagian dataset dengan rasio tersebut mendapatkan akurasi terbaik sebesar 74%.

2.8 Alur Penelitian

Tahapan pertama diawali dengan data sentimen yang telah dikumpulkan melalui media sosial X. Kemudian *dataset* yang telah dikumpulkan akan dilakukan proses *labelling* oleh annotator untuk melihat apakah sentimen yang dilontarkan termasuk ke dalam sentimen positif, negatif, atau netral. Selanjutnya data yang telah dilabeli akan melewati tahap *pre-processing* yang meliputi *cleaning*, *case folding*, *normalization*, *tokenization*, *stopwords*, dan *stemming*. Selanjutnya data yang sudah melewati *pre-processing* akan diolah dalam pembobotan TF-IDF. Setelah melalui tahap pembobotan akan dilakukan tahap klasifikasi dengan metode *Random Forest*. Proses terakhir yang akan dilakukan yaitu dilakukannya proses evaluasi model. Alur proses analisis sentimen menggunakan metode *Random Forest* dapat dilihat pada Gambar 4.



Gambar 4. Flowchart Alur Penelitian

2.9 Crawling Data

Crawling data adalah tahap yang dilakukan dengan tujuan mengumpulkan dan mengunduh informasi dari basis data [22]. Tahapan pertama pada penelitian ini adalah dengan mengumpulkan data pada aplikasi X dengan kata kunci #megathrust. Data tersebut dicrawling dengan menggunakan *library tweet harvest* yang kemudian data akan tersimpan dengan tipe *file csv*.

2.10 Data Labelling

Setelah proses *crawling* data sebanyak 1360 data, selanjutnya akan dilakukan *labelling*. Proses *labelling* ini tidak dilakukan secara otomatis, melainkan oleh pengannotasi yang akan menentukan label setiap data pada tiga kelas, yaitu negatif, positif, serta netral.

2.11 Pre-processing

Pada penelitian [23], dituliskan bahwa *pre-processing* dilakukan untuk memperbaiki struktur data sehingga menjadi lebih rapi dan terorganisir. Sebelum proses klasifikasi sentimen dimulai, data terlebih dulu melalui *pre-processing*. Proses ini bertujuan untuk membersihkan data dengan beberapa tahapan, yaitu:

- a. *Cleaning*
Cleaning merupakan proses untuk membersihkan *noisy* yang ada pada data. *Noisy* tersebut meliputi *link*, *mention*, karakter yang bukan berupa karakter maupun huruf, spasi yang berlebihan, dan *hashtag*.
- b. *Case Folding*
Case Folding merupakan proses yang dilakukan untuk mengubah huruf besar menjadi huruf kecil. Hal ini dilakukan agar mengurangi indeks dengan kata yang sama pada sebuah data.
- c. *Normalization*
Normalization merupakan proses untuk mengubah kata dalam bahasa yang tidak baku menjadi kata yang lebih baku.

- d. *Tokenization*
Tokenization merupakan proses untuk memisahkan kalimat pada suatu data menjadi kata per kata.
- e. *Stopwords*
Stopwords merupakan proses untuk menghilangkan kata yang tidak memiliki pengaruh besar saat akan dilakukan pengklasifikasian. Tujuan dari *stopwords* sendiri adalah untuk melihat seberapa penting kata yang ada pada data.
- f. *Stemming*
Stemming merupakan proses untuk mengubah suatu kata menjadi ke bentuk paling dasarnya.

2.12 Cleaning

Tabel 1. *Cleaning*

Sebelum	Sesudah
#BeritaTerkini Gempa dari Megathrust Jarak 250 Kilometer Dri Pantai Sumbar Diingatkan BMKG https://t.co/TNPIg4vH6C #news #inforiau #infoterkini #riauinfo	Gempa Dari Megathrust Jarak 250 Kilometer Dri Pantai Sumbar Diingatkan BMKG

Pada tabel 1. merupakan contoh proses *cleaning* pada teks. Pada penelitian [24] proses *cleaning* ditujukan untuk menghilangkan informasi yang tidak diperlukan seperti *mention*, tagar atau *hashtag*, dan *link*

2.13 Case Folding

Tabel 2. *Case Folding*

Sebelum	Sesudah
Gempa Dri Megathrust Jarak 250 Kilometer Dari Pantai Sumbar Diingatkan BMKG	gempa dri megathrust jarak 250 kilometer dari pantai sumbar diingatkan bmkkg

Pada tabel 2. merupakan contoh proses *case folding*. Menurut [25] proses *case folding* merupakan poses untuk menyamaratakan huruf atau karakter pada dokumen menjadi kecil.

2.14 Normalization

Tabel 3. *Normalization*

Sebelum	Sesudah
gempa dri megathrust jarak 250 kilometer dari pantai sumbar diingatkan bmkkg	gempa dari megathrust jarak 250 kilometer dari pantai sumbar diingatkan bmkkg

Pada tabel 3. merupakan contoh dari proses *normalization*. Proses *normalization* bertujuan untuk mengubah kata *slang* menjadi kata baku atau formal [26].

2.15 Tokenization

Tabel 4. Tokenization

Sebelum	Sesudah
gempa dari megathrust jarak 250 kilometer dari pantai sumbar diingatkan bmgk	'gempa', 'dari', 'megathrust', 'jarak', '250', 'kilometer', 'dari', 'pantai', 'sumbar', 'diingatkan', 'bmgk'

Pada tabel 4. merupakan contoh dari proses *tokenization*. Proses *tokenization* adalah proses untuk memecah kalimat pada suatu dokumen menjadi kata per kata [27].

2.16 Stopwords

Tabel 5. Stopwords

Sebelum	Sesudah
gempa dari megathrust jarak 250 kilometer dari pantai sumbar diingatkan bmgk	gempa megathrust jarak 250 kilometer pantai sumbar bmgk

Pada tabel 5. merupakan contoh dari proses *stopwords*. Proses ini merupakan tahapan untuk menghapus kata yang tidak terdaftar dalam kriteria *stopwords*.

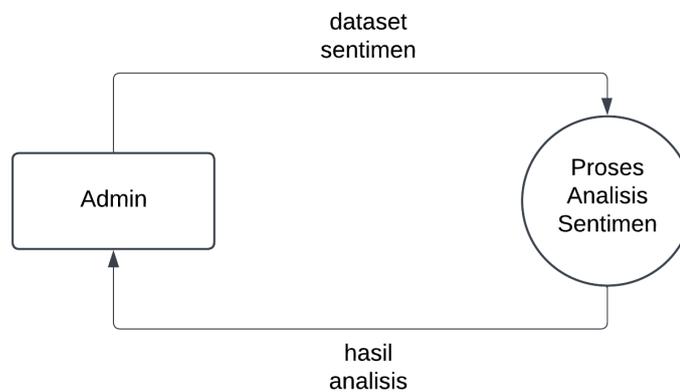
2.17 Stemming

Tabel 6. Stemming

Sebelum	Sesudah
gempa megathrust jarak 250 kilometer pantai sumbar bmgk	gempa megathrust jarak 250 kilometer pantai sumbar bmgk

Pada tabel 6 adalah salah satu implementasi *stemming*, dimana kata yang terdapat pada suatu kalimat akan diubah ke bentuk dasarnya. Pada contoh di atas, kata-kata yang terdapat di pada teks sudah dalam bentuk dasar, sehingga tidak ada perubahan yang terjadi.

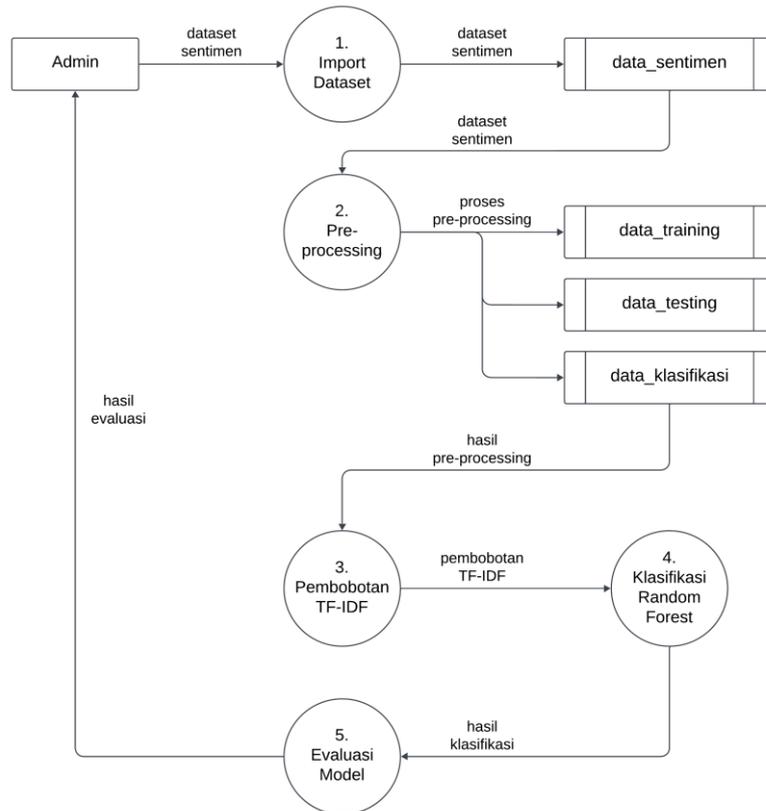
2.18 DFD Level 0



Gambar 5. DFD Level 0

Gambar 5. merupakan alur data diagram *level 0* aplikasi. Pengguna yang telah memiliki *dataset* sentimen akan melakukan proses analisis sentimen. Kemudian hasil dari analisis tersebut akan diberikan kembali kepada pengguna.

2.19 DFD Level 1



Gambar 6. DFD Level 1

Gambar 6 di atas merupakan alur data diagram *level 1* pada aplikasi analisis sentimen. Pertama pengguna akan melakukan *import dataset* dengan *file* bertipe *csv*. Setelah itu, data yang telah diimpor akan disimpan dalam *database* pada tabel *data_sentimen*. Setelah *dataset* tersimpan, tahapan *pre-processing* akan dilakukan untuk membersihkan data sentimen. Data yang telah melalui tahap *pre-processing* akan disimpan dan dibagi ke dalam tiga tabel yaitu *data_training*, *data_testing*, dan *data_klasifikasi*. Ketika tahap *pre-processing* sudah selesai, data pada tabel *data_klasifikasi* akan digunakan untuk proses pembobotan dengan TF-IDF. Apabila proses pembobotan sudah selesai, akan dilakukan klasifikasi dengan metode *Random Forest* dan hasil klasifikasi tersebut akan dievaluasi.

3. HASIL DAN PEMBAHASAN

3.1 Halaman *Dashboard*

Pada halaman ini pengguna akan dapat melihat berapa banyak data yang digunakan dalam proses analisis sentimen, baik jumlah data *training*, data *testing*, dan total data. Tidak hanya dalam bentuk angka, namun *data training* dan *data testing* juga divisualisasikan dengan menggunakan diagram batang. Hasil tampilan *dashboard* dari aplikasi analisis sentimen berbasis *website* dapat dilihat pada gambar 7.



Gambar 7. Halaman *Dashboard*

3.2 Halaman *Data Sentimen*

#	Username	Text
1	TsaabitahDAA	Perhatian!!! Ini ttg POTENSI GEMPA BESAR MEGATHRUST dan ANCAMAN TSUNAMI SETINGGI 20 METER DI PULAU JAWA LOHHH https://t.co/iqIMq9WmGU
2	aoxyabcd	nah kann megathust lagi :) sejak pindah ke Depok jadi khawatir sama megathrust.
3	winardonol	@RANGERmounts Megathrust itu brp SR Kang ?
4	RadioElshinta	Mewaspadai ancaman gempa besar dari patahan megathrust di Selat Sunda mencapai M 8 7 #KomisiAndaElshinta https://t.co/MnzJOshhNj

Gambar 8. Halaman *Data Sentimen*

Pada gambar 8 merupakan tampilan dari halaman data sentimen. Di halaman ini super admin dan admin dapat melakukan *import* data dalam bentuk *file* csv. Kemudian data akan menampilkan *username* dan *text* cuitan pada tabel yang sudah tersedia. Ada pula tombol untuk menghapus data sentimen yang sudah di-*import*kan jika pengguna ingin mengganti *dataset* yang ada.

3.3 Halaman *Pre-processing Data Training*

Pada halaman ini, terdapat tombol untuk melakukan tahapan *pre-processing* pada data. Data yang sudah mengalami tahap *pre-processing* akan ditampilkan pada tabel dengan kolom sebelah kiri untuk melihat *text* mentah dan kolom sebelah kanan untuk melihat *text* yang sudah diproses. Hasil tampilan dari halaman *pre-processing* data *training* dapat dilihat pada Gambar 9.



#	Text	Hasil Pre-processing
1	@MeidaaMei29 @Tan_Mar3M c. Dr ujung Sumatera sampe Indonesia timur rentan gempa krn ada sesar memanjang (megathrust) yg kapan aja bisa terjadi. Kalo itu terjadi.. lo mau NKRI lumpuh. So borneo much save then others. 2. Orang pilih makan drpd internet.. notabene manfaat nya jelas	sih ujung Sumatera Indonesia timur rentan gempa sesar panjang megathrust kalo lo nkri lumpuh so borneo much save then others 2 orang pilih makan internet notabene manfaat nya
2	GEMPA YAALLAH LGSG NEGTHINK KE MEGATHRUST	gempa yaallah langsung negthink megathrust
3	Masyarakat Diminta Tidak Panik Terkait Potensi Gempa Megathrust #merahputihcom #gempa #megathrust https://t.co/5tQjldZmeO	masyarakat panik kait potensi gempa megathrust

Gambar 9. Halaman *Pre-processing Data Training*

3.4 Halaman *Klasifikasi Data Training*

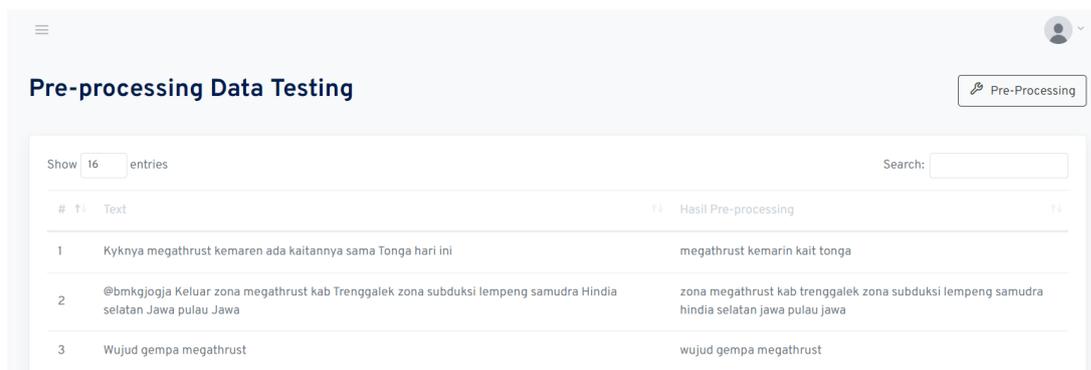


#	Text	Label	Hasil Pre-processing
1	@MeidaaMei29 @Tan_Mar3M c. Dr ujung Sumatera sampe Indonesia timur rentan gempa krn ada sesar memanjang (megathrust) yg kapan aja bisa terjadi. Kalo itu terjadi.. lo mau NKRI lumpuh. So borneo much save then others. 2. Orang pilih makan drpd internet.. notabene manfaat nya jelas	positif	sih ujung Sumatera Indonesia timur rentan gempa sesar panjang megathrust kalo lo nkri lumpuh so borneo much save then others 2 orang pilih makan internet notabene manfaat nya
2	GEMPA YAALLAH LGSG NEGTHINK KE MEGATHRUST	positif	gempa yaallah langsung negthink megathrust
3	Masyarakat Diminta Tidak Panik Terkait Potensi Gempa Megathrust #merahputihcom #gempa #megathrust https://t.co/5tQjldZmeO	netral	masyarakat panik kait potensi gempa megathrust

Gambar 10. Halaman *Klasifikasi Data Training*

Gambar 10 merupakan hasil tampilan dari halaman klasifikasi data *training*. Pada halaman ini, pengguna dapat melihat hasil klasifikasi yang telah dilakukan oleh annotator pada *data training*.

3.5 Halaman *Pre-processing Data Testing*

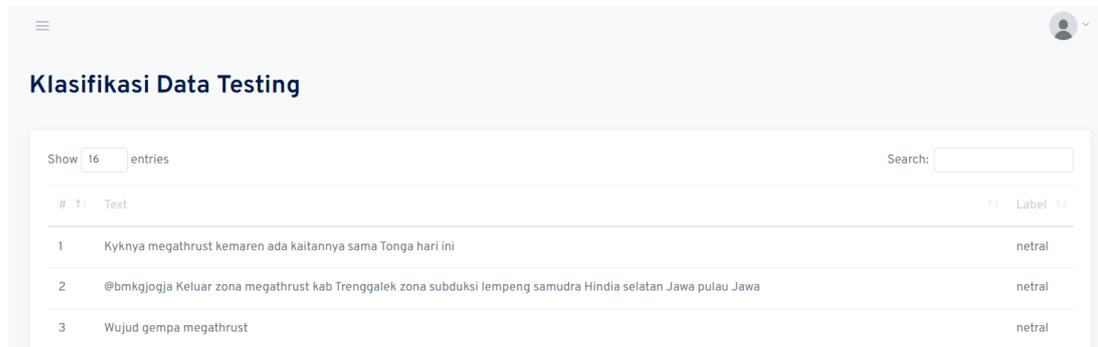


#	Text	Hasil Pre-processing
1	Kyknya megathrust kemaren ada kaitannya sama Tonga hari ini	megathrust kemarin kait tonga
2	@bmgjogja Keluar zona megathrust kab Trenggalek zona subduksi lempeng samudra Hindia selatan Jawa pulau Jawa	zona megathrust kab trenggalek zona subduksi lempeng samudra hindia selatan Jawa pulau Jawa
3	Wujud gempa megathrust	wujud gempa megathrust

Gambar 11. Halaman *Pre-processing Data Testing*

Pada gambar 11 merupakan tampilan dari halaman *pre-processing data testing*. Pada halaman ini, terdapat tombol untuk melakukan tahapan *pre-processing* pada data. Sama halnya pada halaman *pre-processing* pada data *training*, data *testing* yang sudah mengalami tahap *pre-processing* akan ditampilkan pada tabel dengan kolom sebelah kiri untuk melihat *text* mentah dan kolom sebelah kanan untuk melihat *text* yang sudah diproses.

3.6 Halaman Klasifikasi Data Testing

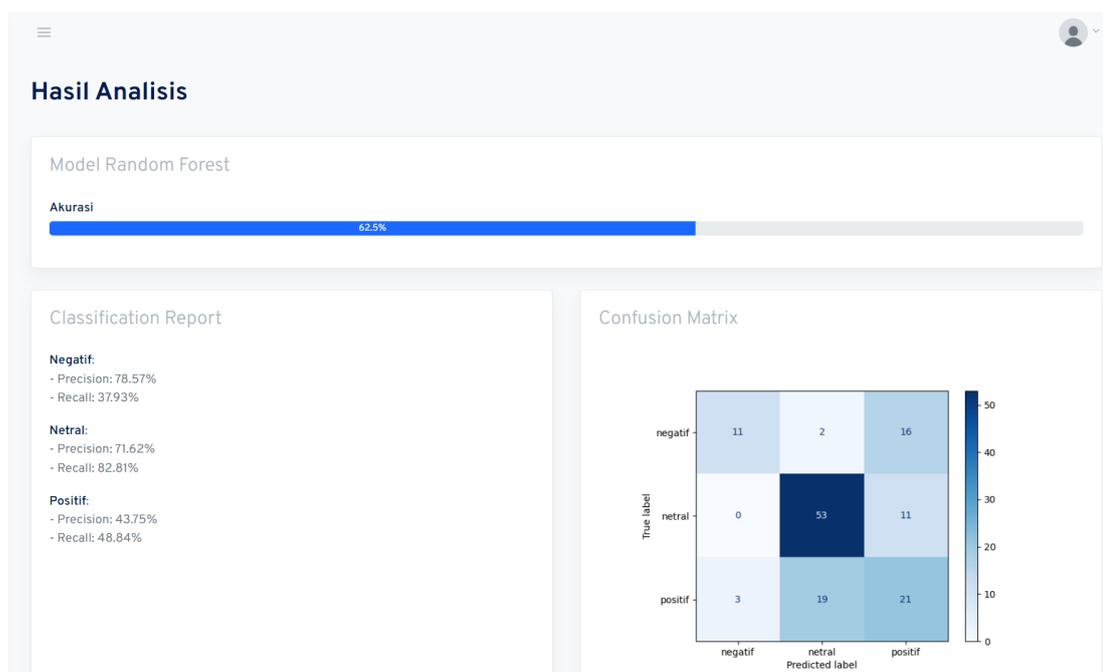


#	Text	Label
1	Kyknya megathrust kemaren ada kaitannya sama Tonga hari ini	netral
2	@bmkjogja Keluar zona megathrust kab Trenggalek zona subduksi lempeng samudra Hindia selatan Jawa pulau Jawa	netral
3	Wujud gempa megathrust	netral

Gambar 12. Halaman Klasifikasi Data Testing

Pada gambar 12. merupakan tampilan klasifikasi data *testing*. Pada halaman ini, pengguna dapat melihat hasil klasifikasi yang telah dilakukan oleh anotator pada data *testing*.

3.7 Halaman Hasil Analisis

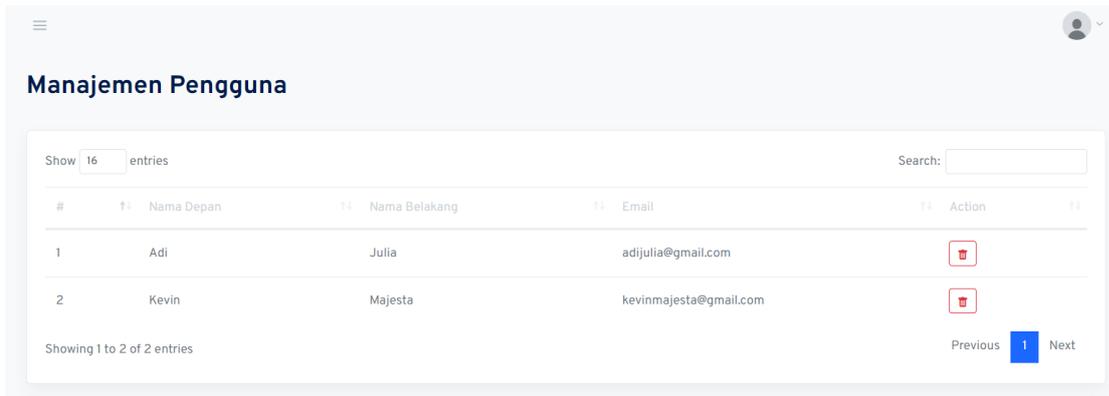


Gambar 13. Halaman Hasil Analisis

Pada gambar 13. merupakan tampilan hasil analisis. Pada halaman ini pengguna dapat melihat seberapa baik akurasi model dalam melakukan klasifikasi sentimen. Terdapat juga *classification report*, *confusion matrix*, dan terakhir *wordcloud* untuk melihat kata yang sering dikeluarkan baik pada sentimen positif, negatif, maupun netral.

3.8 Halaman Manajemen User

Pada halaman ini *super admin* dapat melihat dan menghapus data *user* yang sudah melakukan registrasi pada aplikasi. *User* yang sudah teregistrasi akan ditampilkan pada tabel, serta *super admin* dapat menghapus *user* yang sudah terdaftar. Hasil tampilan halaman manajemen *user* dapat dilihat pada gambar 14.



Gambar 14. Halaman Manajemen User

3.9 Evaluasi Model

Untuk melakukan pengujian model *Random Forest* dalam melakukan pengklasifikasian, *dataset* dibagi dengan menggunakan proporsi 90:10. Pembagian dilakukan dengan jumlah *data training* sebesar 90% dari total data dan jumlah data *testing* sebesar 10% dari total data keseluruhan yaitu sebanyak 1360 data. Evaluasi dilakukan dengan menghitung hasil klasifikasi berdasarkan tabel *confusion matrix* yang merepresentasikan perbandingan antara label yang dilakukan oleh sistem dengan label yang dihasilkan oleh model. Tabel *confusion matrix* dapat dilihat pada Tabel 7.

Tabel 7. Perhitungan *Confusion Matrix*

True Class	Predicted Class			Total
	Negatif	Netral	Positif	
Negatif	11	4	14	29
Netral	0	54	10	64
Positif	4	18	21	43
Total	15	76	45	136

Setelah dilakukan perhitungan dengan *confusion matrix*, selanjutnya nilai yang telah didapatkan akan digunakan untuk menghitung *accuracy* pada persamaan (4), *precision* pada persamaan (5), dan *recall* pada persamaan (6). Berikut adalah hasil dari perhitungan *confusion matrix*:

$$Accuracy = \frac{TP_{Positif} + TP_{Negatif} + TP_{Netral}}{Total\ Data} \times 100\% = \frac{11 + 54 + 21}{136} \times 100\% = 63.24\% \quad (7)$$

$$Precision_{Netral} = \frac{TP_{Netral}}{TP_{Netral} + FP_{Netral}} \times 100\% = \frac{54}{54 + 22} \times 100\% = 71.05\% \quad (8)$$

$$Recall_{Netral} = \frac{TP_{Netral}}{TP_{Netral} + FN_{Netral}} \times 100\% = \frac{54}{54 + 10} \times 100\% = 84.38\% \quad (9)$$

Pada persamaan di atas, didapatkan *accuracy* model sebesar 63.24%. Beberapa faktor yang dapat memengaruhi rendahnya *accuracy* salah satunya adalah data yang digunakan memiliki kelas yang tidak seimbang, sehingga sangat disarankan pada penelitian selanjutnya dilakukan penyeimbangan kelas pada *dataset* yang ada.

3.10 Pengujian Sistem

Adanya pengujian sistem ditujukan memahami bagaimana cara sistem bekerja secara menyeluruh untuk mencapai target yang telah ditentukan. Melalui penerapan pengujian *blackbox*, peneliti dapat mengevaluasi bagaimana suatu sistem bekerja sesuai dengan yang diharapkan. Metode ini menekankan penilaian berdasarkan kemampuan fungsionalitas. Hasil dari pengujian dapat dilihat di Tabel 8.

Tabel 8. Pengujian Sistem

No	Skenario Pengujian	Hasil yang Diharapkan	Hasil Uji
1	<i>Super admin, admin, dan user</i> dapat memasukkan <i>username</i> dan <i>password</i>	Pengguna berhasil mengakses sistem setelah <i>login</i>	Sesuai
2	<i>Super admin, admin, dan user</i> dapat melakukan <i>logout</i>	Pengguna dapat <i>logout</i> dari aplikasi	Sesuai
3	<i>User</i> dapat melakukan registrasi akun	Akun baru berhasil dibuat oleh pengguna	Sesuai
4	<i>Super admin</i> dan <i>admin</i> dapat melakukan <i>import</i> dan hapus <i>dataset</i>	Data berhasil ditambahkan dan dihapus sesuai kebutuhan	Sesuai
5	<i>Super admin</i> dan <i>admin</i> dapat melakukan <i>pre-processing</i>	Proses <i>pre-processing</i> berjalan tanpa kendala	Sesuai
6	<i>Super admin, admin, dan user</i> dapat melihat hasil analisis	Hasil analisis dapat ditampilkan di antarmuka	Sesuai
7	<i>Super admin</i> dapat melakukan manajemen <i>user</i>	<i>Super admin</i> dapat mengelola data <i>user</i> dengan sukses	Sesuai

Berdasarkan Tabel 8. terdapat 7 skenario pengujian. Pada baris pertama hingga keempat merupakan pengujian autentikasi pada aplikasi. Diharapkan pengguna yang menggunakan aplikasi ini dapat melakukan *login*, *logout*, dan registrasi akun. Selanjutnya adalah fitur-fitur yang terdapat di dalam aplikasi, yaitu adanya fitur *import* dan hapus *dataset* yang dapat dilakukan oleh *super admin* dan *admin*. Kemudian terdapat fitur untuk melakukan *pre-processing* untuk membersihkan data sentimen yang dapat dilakukan oleh *super admin* dan *admin*. Setelah itu terdapat fitur untuk melihat hasil analisis yang telah dilakukan oleh model, fitur ini dapat dilihat oleh *super admin, admin, dan user*. Terakhir adalah adanya fitur manajemen *user* yang hanya dapat diakses oleh *super admin*.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, didapatkan bahwa sistem dapat mengidentifikasi kelas netral dengan cukup baik. Dengan menggunakan data sentimen sebanyak 136 sebagai pengujian, didapatkan tingkat *accuracy* sebesar 63,24%, *precision* 71,05%, dan *recall* mencapai 84,38%. Tingkat akurasi ini juga dipengaruhi oleh seberapa baik data yang didapatkan, jumlah data, dan seberapa baik tahapan *pre-processing* dilakukan. Kontribusi utama penelitian ini adalah penerapan analisis sentimen untuk membantu lembaga maupun komunitas sadar bencana dalam memahami persepsi masyarakat, sehingga mendukung strategi komunikasi dan mitigasi yang lebih efektif. Pada penelitian selanjutnya, diharapkan dapat dilakukan penyeimbangan data, di antaranya dengan menggunakan metode SMOTE. Selain itu, diharapkan adanya penggunaan metode lain, seperti SVM, untuk membandingkan hasil yang didapatkan.

DAFTAR PUSTAKA

- [1] W. Kurniawan, D. Daryono, I. Kerta, and T. Triwinugroho, "Analisis Sistem Peringatan Dini Tsunami di Zona Megathrust Selat Sunda Guna Mewujudkan Ketahanan Nasional," *PENDIPA J. Sci. Educ.*, vol. 6, no. 2, pp. 457–464, 2022, doi: 10.33369/pendipa.6.2.457-464.
- [2] A. Fadilah, "Tentang Gempa di Selat Sunda dan Mentawai-Siberut yang 'Tinggal Menunggu Waktu,'" BMKG. [Online]. Available: <https://www.bmkg.go.id/berita/?p=tentang-gempa-di-selat-sunda-dan-mentawai-siberut-yang-tinggal-menunggu-waktu&lang=ID&s=detil>
- [3] T. Arbar, "BMKG Bilang Gempa Megathrust RI Hanya Tunggu Waktu, Ini Zona Merahnya," CNBC Indonesia. [Online]. Available: <https://www.cnbcindonesia.com/tech/20240825063902-37-566129/bmkg-bilang-gempa-megathrust-ri-hanya-tunggu-waktu-ini-zona-merahnya>
- [4] F. Melinda and Suhartono, "Antisipasi ancaman bencana gempa megathrust 16," 2024.
- [5] R. S. Yuliatmoko, Y. H. Perdana, and A. A. Martha, "Distribusi Frekuensi Gempabumi Dan Dimensi Fraktal Pada Seismik Gap Di Indonesia Earthquake Frequency-Magnitude Distribution and Fractal Dimension on Seismic Gap in Indonesia," *J. Meteorol. Dan Geofis.*, vol. 22, no. 2, pp. 55–66, 2021.

- [6] A. Permana, "Kepanikan Netizen Usai Estimasi Kekuatan Gempa Megathrust di Indonesia, Warganet: Inpo Tanah Murah di Kalimantan," Warnanusa. [Online]. Available: <https://www.warnanusa.com/nasional/83513346221/kepanikan-netizen-usai-estimasi-kekuatan-gempa-megathrust-di-indonesia-warganet-inpo-tanah-murah-di-kalimantan>
- [7] G. N. Purba, "Tanggapan Masyarakat soal Gempa Megathrust," Metrotvnews. [Online]. Available: <https://www.metrotvnews.com/play/bmRCeDXX-tanggapan-masyarakat-soal-gempa-megathrust>
- [8] D. Ayu, "Tanggapan BPBD Kota Batu Soal Megathrust, Minta Masyarakat Tak Khawatir Berlebihan," TribunBatu. [Online]. Available: <https://jatim.tribunnews.com/2024/09/05/tanggapan-bpbd-kota-batu-soal-megathrust-minta-masyarakat-tak-khawatir-berlebihan>
- [9] A. Ramadhan, B. Susetyo, and Indahwati, "Penerapan Metode Klasifikasi Random Forest Dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan," *J. Pendidik. dan Kebud.*, vol. 4, no. 2, pp. 169–182, 2019, doi: 10.24832/jpnk.v4i2.1327.
- [10] M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.
- [11] T. A. U. Azmi, L. Hakim, D. C. R. Novitasari, and W. D. U. D. Utami, "Application Random Forest Method for Sentiment Analysis in Jamsostek Mobile Review," *Telematika*, vol. 20, no. 1, p. 117, 2023, doi: 10.31315/telematika.v20i1.8868.
- [12] F. I. Kurniadi and P. D. Larasati, "Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 67–72, 2022, doi: 10.47970/siskom-kb.v6i1.328.
- [13] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.
- [14] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealis.v6i2.3036.
- [15] Y. Romadhoni and K. F. H. Holle, "Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM," *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 118–124, 2022, doi: 10.30591/jpit.v7i2.3191.
- [16] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks," *SINTESIA J. Sist. dan Teknol. Inf. Indones.*, vol. 1, no. 2, pp. 81–88, 2023.
- [17] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naive Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
- [18] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [19] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [20] D. Nurwahidah, G. Dwilestari, N. Dienwati Nuris, and R. Narasati, "Analisis Sentimen Data Ulasan Pengguna Aplikasi Google Kelas Pada Google Play Store Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3673–3678, 2024, doi: 10.36040/jati.v7i6.8245.
- [21] D. Manuel, Y. Sinurat, D. E. Ratnawati, and D. W. Brata, "Analisis Sentimen Terhadap Kenaikan Cukai Rokok pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 17–25, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [22] R. Fiddiyansyah, Izra Noor Zahara Aliya, and Moh Azzam Priyanto, "Dampak Identity Theft Berdasarkan Artikel Berita Dan Crawling Data Sentimen Twitter," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 629–638, 2023, doi: 10.33005/sitasi.v3i1.399.
- [23] D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021, doi: 10.25126/justsi.v2i1.24.
- [24] T. S. Sabrila, Y. Azhar, and C. S. K. Aditya, "Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 10–19, 2022, doi: 10.14421/jiska.2022.7.1.10-19.

- [25] M. R. D. Fachreza, S. Suhartono, and M. A. Yaqin, "Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 3, pp. 243–251, 2023, doi: 10.14421/jjska.2023.8.3.243-251.
- [26] P. Ayuningtyas, S. Khomsah, T. Informatika, F. Informatika, S. Data, and F. Informatika, "Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU," vol. 9, no. 3, pp. 217–229, 2024.
- [27] A. Hendra and F. Fitriyani, "Analisis Sentimen Review Halodoc Menggunakan Naïve Bayes Classifier," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 78–89, 2021, doi: 10.14421/jjska.2021.6.2.78-89.