

PEMETAAN SENTIMEN MEDIA SOSIAL TERHADAP ISU  
MEGATHRUST DI SELATAN PULAU JAWA MENGGUNAKAN  
*NATURAL LANGUAGE PROCESSING*

Bayu Kurniawan

21/480093/GE/09662

INTISARI

Pulau Jawa memiliki kerawanan seismik yang tinggi akibat keberadaan zona *megathrust* di sisi selatannya yang memiliki celah seismik (*seismic gap*) dengan potensi tsunami hingga 20 meter. Meskipun frekuensi terjadinya rendah, dampak destruktifnya sangat besar. Potensi bencana ini memicu perhatian dan diskursus yang luas di kalangan masyarakat khususnya pengguna media sosial. Tingginya konsentrasi pengguna internet di Pulau Jawa yang mencapai 54,96% menjadikan media sosial sebagai wadah utama penyampaian opini publik. Berdasarkan hal ini, diperlukan pemetaan sentimen publik terkait isu *megathrust* untuk melihat respon masyarakat lebih spesifik pada lokasi geografis yang berbeda. Penelitian ini berhasil membangun basis data sentimen spasial dari 4.193 posting Twitter (X) yang diseleksi menjadi 325 data yang mencakup 54 dari 119 kabupaten/kota di Pulau Jawa. Klasifikasi sentimen dilakukan menggunakan model IndoBERT yang memiliki performa lebih baik dengan akurasi 85% dibandingkan Microsoft Azure (53%). Hasil klasifikasi didominasi oleh sentimen negatif (65,3%), diikuti oleh netral (32,1%), dan positif (2,6%), mencerminkan kecemasan dan kekhawatiran yang meluas di kalangan masyarakat. Visualisasi spasial memperlihatkan bahwa sentimen negatif mendominasi sebagian besar wilayah di Pulau Jawa. Namun, *Geographically Weighted Regression* (GWR) yang diterapkan pada wilayah Jabodetabek dan Jawa Barat menunjukkan kemampuan penjelas yang sangat terbatas. Nilai  $R^2$  model hanya 0,221, dan tidak ada dari variabel geografis atau sosio-ekonomi yang diuji (indeks risiko, frekuensi gempa bumi, tingkat pengangguran, atau tahun pendidikan yang diharapkan) yang ditemukan secara statistik signifikan dalam menjelaskan variasi sentimen. Hal ini disebabkan oleh faktor seperti penyebaran informasi digital, narasi media pembaca, dan kesenjangan data akibat medan pegunungan kasar yang menghambat penetrasi masyarakat ke media sosial.

**Kata kunci:** *megathrust*, analisis sentimen, IndoBERT, Twitter (X), *Geographically Weighted Regression* (GWR).

## MAPPING SOCIAL MEDIA SENTIMENT TOWARDS THE MEGATHRUST ISSUE IN SOUTHERN JAVA USING NATURAL LANGUAGE PROCESSING

Bayu Kurniawan

21/480093/GE/09662

### ABSTRACT

Java Island exhibits high seismic vulnerability due to the presence of a megathrust zone on its southern coast, which features a seismic gap with a tsunami potential of up to 20 meters. Despite its low frequency of occurrence, the potential destructive impact is immense. The potential for this disaster has triggered widespread attention and discourse among the public, especially social media users. The high concentration of internet users in Java, reaching 54.96%, establishes social media as a primary platform for public opinion. Consequently, mapping public sentiment regarding the megathrust issue is essential to discern specific societal responses across different geographic locations. This study successfully constructed a spatial sentiment database from 4,193 Twitter (X) posts, curated to 325 data points covering 54 of the 119 regencies/cities in Java. Sentiment classification was conducted using the IndoBERT model, which outperformed Microsoft Azure with an accuracy of 85% compared to 53%. The classification results were dominated by negative sentiment (65.3%), followed by neutral (32.1%) and positive (2.6%), reflecting widespread anxiety and concern among the public. Spatial visualization indicates that negative sentiment dominates most regions in Java. However, Geographically Weighted Regression (GWR) applied to the Jabodetabek and West Java regions demonstrated very limited explanatory power. The model's  $R^2$  value was only 0.221, and none of the tested geographic or socio-economic variables (risk index, earthquake frequency, unemployment rate, or expected years of schooling) were found to be statistically significant in explaining sentiment variation. This is attributed to factors such as digital information dissemination, media narratives, and data gaps resulting from rugged mountainous terrain hindering social media penetration.

**Keywords:** megathrust, sentiment analysis, IndoBERT, Twitter(X), Geographically Weighted Regression (GWR)