

## Analisis Sentimen Data Gempa Bumi Menggunakan Support Vector Machine

Irfani Firdausy<sup>1</sup>, Fauziyah Amini<sup>2</sup>, Ni'mah Firsta Cahya Susilo<sup>3</sup>, Moch. Habil Noer<sup>4</sup>, Tri Gunantohadi<sup>5</sup>

Jurusan Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang

Email: [1200605210001@student.uin-malang.ac.id](mailto:1200605210001@student.uin-malang.ac.id), [2200605210002@student.uin-malang.ac.id](mailto:2200605210002@student.uin-malang.ac.id),  
[3200605210003@student.uin-malang.ac.id](mailto:3200605210003@student.uin-malang.ac.id), [4200605210004@student.uin-malang.ac.id](mailto:4200605210004@student.uin-malang.ac.id)  
[5200605210005@student.uin-malang.ac.id](mailto:5200605210005@student.uin-malang.ac.id)

### Abstrak

Indonesia merupakan negara maritim yang terletak pada 3 lempeng dunia atau biasa disebut Ring of Fire yang sering menimbulkan gempa bumi. Gempa bumi merupakan ancaman terbesar yang dihadapi oleh potensi tsunami di dalamnya yang dapat menimbulkan kerusakan bahkan menimbulkan korban jiwa. Malang merupakan daerah di Jawa Timur yang berada pada daerah pegunungan yang memiliki gunung aktif. Dengan pesatnya penggunaan social media dalam kehidupan sehari-hari dapat memberikan berbagai macam informasi yang bisa disampaikan, salah satunya media social Twitter. Teks mining dapat mengeksplorasi data gempa yang ada pada Twitter dengan tagar GempaBumiMalang. Analisis Sentimen adalah salah satu cara yang digunakan untuk mengetahui penilaian atau opini pengguna Twitter. Berdasarkan hasil dan pembahasan dapat ditarik kesimpulan dengan menerapkan analisis sentimen dapat diketahui akurasi sebesar 63.5%. Hasil true negative ditemukan sebesar 26, false negative sebesar 4, false positive sebesar 69, dan true positive ditemukan sebesar 101. Dari Gambar 2, diperoleh hasil sensitivity sebesar 96,19%, specificity sebesar 27.37%, negative predictive value sebesar 86.67, dan precision sebesar 59.41%

**Keywords:** *analisis sentimen, gempa bumi, support vector machine, confusion matrix*

### Abstract

Indonesia is a maritime country located on 3 plates of the world or commonly called the Ring of Fire which often causes earthquakes. Earthquakes are the biggest threat faced by a potential tsunami in it which can cause damage and even cause casualties. Malang is an area in East Java which is located in a mountainous area that has an active volcano. With the rapid use of social media in everyday life, it can provide various kinds of information that can be conveyed, one of which is social media Twitter. Text mining can explore earthquake data on Twitter with the hashtag BumiMalang Earthquake. Sentiment analysis is one of the methods used to find out the ratings or opinions of Twitter users. Based on the results and discussion, it can be concluded that by applying sentiment analysis, it can be seen that the accuracy is 63.5%. True negative results were found to be 26, false negatives were 4, false positives were 69, and true positives were 101. From Figure 2, the sensitivity results were 96.19%, specificity was 27.37%, negative predictive value was 86.67, and precision by 59.41%

**Keywords:** *sentiment analysis, earthquake, support vector machine, confusion matrix*

### 1. PENDAHULUAN

Indonesia merupakan negara kepulauan yang terdiri dari lebih dari tujuh belas ribu pulau, termasuk Sumatra, Jawa, Sulawesi, dan sebagian Kalimantan dan New Guinea. Indonesia adalah

negara kepulauan terbesar di dunia dan negara terluas ke-14 berdasarkan luas daratan, dengan luas 1.904.569 kilometer persegi (735.358 mil persegi). Secara tektonik, Indonesia sangat tidak stabil, hal ini disebabkan oleh Indonesia yang terletak pada *Ring of Fire* yang membuat

Indonesia menjadi lokasi banyak gunung berapi dan sering terjadi gempa. Gempa bumi merupakan ancaman terbesar yang dihadapi oleh potensi tsunami di dalamnya yang dapat menimbulkan kerusakan bahkan menimbulkan korban jiwa.

Malang merupakan daerah yang dikelilingi oleh pegunungan dan jajaran pegunungan. Kota ini dikelilingi oleh Gunung Arjuno di utara; Gunung Semeru di sebelah timur; Gunung Kawi dan Gunung Butak di barat; Gunung Kelud di selatan. Gunung berapi aktif yang populer Gunung Bromo terletak sekitar 25 km (16 mil) timur kota.

Berkembangnya teknologi saat ini mempermudah manusia untuk memperoleh berbagai informasi secara cepat terutama dengan maraknya informasi yang diperoleh dari media sosial. *Text mining* adalah proses memeriksa koleksi dokumen yang besar untuk menemukan informasi baru atau membantu menjawab pertanyaan penelitian tertentu. Data terstruktur yang dibuat oleh text mining dapat diintegrasikan ke dalam basis data, gudang data, atau dasbor intelijen bisnis dan digunakan untuk analisis deskriptif, preskriptif, atau prediktif.

Analisis sentimen adalah proses mendeteksi sentimen positif atau negatif dalam teks. Ini sering digunakan oleh bisnis untuk mendeteksi sentimen dalam data sosial, mengukur reputasi merek, dan memahami pelanggan. Analisis sentimen dapat mengidentifikasi isu-isu kritis secara real-time, misalnya krisis PR di media sosial meningkat? Apakah pelanggan yang marah akan melakukan churn? Model analisis sentimen dapat membantu untuk segera mengidentifikasi situasi semacam ini, sehingga dapat segera mengambil tindakan.

Penelitian ini menggunakan prediksi pada data yang telah diperoleh. Selain melakukan klasifikasi linier, SVM dapat secara efisien melakukan klasifikasi non-linier menggunakan apa yang disebut trik kernel, secara implisit memetakan inputnya ke dalam ruang fitur berdimensi tinggi. SVM sangat membantu dalam kategorisasi teks dan hypertext, karena aplikasinya dapat secara signifikan mengurangi kebutuhan akan instance pelatihan berlabel baik dalam pengaturan induktif dan transduktif standar. Beberapa metode untuk parsing semantik dangkal didasarkan pada mesin vektor dukungan (Pradhan et al, 2004)

Pada penelitian ini, kami akan melakukan

analisa sentimen data dari cuitan pengguna twitter yang berhubungan dengan gempa bumi. Data yang diperoleh akan diolah dengan menggunakan metode *support vector machine* untuk prediksi “positif” dan “negatif”.

## 2. LITERATURE REVIEW

Analisis sentimen adalah metode untuk menganalisis sebagian data untuk mengetahui emosi manusia. Analisis sentimen dapat dikategorikan kedalam tiga task, yaitu *informative text detection*, *information extraction* dan *sentiment interestingness classification* (*emotional*, *polarity identification*). Sentiment classification (negatif atau positif) digunakan untuk memprediksi sentiment polarity berdasarkan data sentimen dari pengguna (Pan, Ni, Sun, Yang, & Chen, 2010). Textual sentiment analysis telah banyak digunakan, penggunaannya tidak sebatas dalam area penelitian ilmiah tetapi juga untuk kebutuhan business marketing dan teknologi (Chintala, 2012).

Menurut Go (2009), analisis sentimen adalah sebuah area penelitian yang menonjol dan berkembang aktif yang dipengaruhi oleh pertumbuhan teknologi media sosial yang cepat. Melalui media sosial terdapat peluang untuk mengakses opini dari sejumlah orang pada berbagai jenis bisnis, isu dunia dan isu sosial (Go, Huang, & Bhayani, 2009). Pada jurnal “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional” (Jurnal TEKNO KOMPAK, Vol. 15, No. 1, P-ISSN: 1412-9663, E-ISSN : 2656-3525, Hal. 131-145) sentimen analisis adalah mengumpulkan data opini melalui media data online. Kamus data opini yang dikumpulkan diklasifikasi menjadi kamus opini negatif dan positif. Untuk melakukan klarifikasi peneliti menggunakan bantuan secara online maupun offline untuk melakukan klarifikasi agar data kamus sesuai dan meminimalkan subjektivitas.

Pada penelitian “Sentiment analysis for the tweets that contain the word “earthquake” menunjukkan bahwa tweet yang mengandung kata “gempa bumi” mengungkapkan perasaan negatif di lebih dari 80% kasus, perhitungan dilakukan dengan berbagai metode (Pirnaui, 2018). Pada saat yang sama, untuk mengidentifikasi sentimen pengguna mengenai situasi tertentu – gempa bumi, dalam hal ini,

sangat berguna untuk mengetahui daftar kata-kata yang berdekatan, yang sering digunakan dalam diskusi tentang topik ini. Penelitian tersebut mengidentifikasi daftar kata-kata yang dapat ditemukan dalam tweet yang dikumpulkan. Dari semua kata ini, hanya kata-kata dengan lebih dari 25% kemunculan dalam tweet yang dikumpulkan dalam situasi gempa yang diekstraksi. Mengetahui daftar kata kunci ketika menafsirkan data yang dikumpulkan dari Jejaring Sosial, dalam situasi gempa bumi, berkontribusi untuk mengurangi waktu yang dibutuhkan untuk mengidentifikasi pendapat dan pengambilan keputusan.

Penelitian Vo dan Collier (2013) melakukan analisa pelacakan emosi dalam gempa bumi di Jepang untuk mengelola situasi dengan lebih baik. Untuk mencapai tujuan tersebut, diusulkan emosi yang tepat untuk situasi gempa. Mereka juga mengusulkan klasifikasi tweet Peduli/Tidak Peduli gempa dan klasifikasi tweet emosi yang sama sekali berbeda dari karya yang tersedia terkait gempa. Fitur n-gram sederhana adalah pilihan terbaik untuk mengklasifikasikan bahasa Jepang yang aglutinatif dan bahasa Twitter yang berisik. Emosi dalam interval waktu tanggal gempa mengungkapkan wawasan pengguna Twitter selama masa sulit seperti itu

### 3. METODOLOGI

Dalam penelitian ini kami melakukan beberapa tahap, yaitu :

#### 1. Pengumpulan Data

Data pada penelitian ini kami peroleh dari cuitan pengguna twitter yang berkenaan dengan gempa bumi. Data yang diperoleh sejumlah 502 cuitan twitter dengan tagar #GempaBumiMalang.

#### 2. Preprocessing

*Preprocessing* adalah tahap menyiapkan data agar data layak untuk digunakan pada proses selanjutnya, berikut ini adalah beberapa tahap pada preprocessing:

##### a. Tokenisasi

Tokenisasi adalah proses untuk memisahkan atau memenggal kalimat kedalam beberapa kata.

##### b. Stemming

Proses stemming adalah menghilangkan imbuhan pada awalan dan akhiran. Proses stemming akan menghasilkan kata dasar dari sebuah kata. Dimana dari hasil proses tersebut akan didapatkan sebuah informasi mengenai banyaknya term yang muncul dalam sebuah dokumen setelah dilakukan perhitungan term

frequency.

#### c. Stopword Removal

Pada tahap ini dilakukan penghapusan kata-kata yang tidak deskriptif (tidak penting).

#### 3. Prediksi

Dalam proses prediksi ini, setiap kata akan di kalsifikasi menjadi sentiment negative dan sentiment positif.

#### 4. Support Vector Machine

Dalam *machine learning*, support vector machine (SVM) adalah *supervised learning* dengan algoritme pembelajaran terkait yang menganalisis data untuk analisis klasifikasi dan regresi. Dalam algoritma SVM, margin antara titik data dan hyperplane akan dimaksimalkan. Fungsi kerugian yang membantu memaksimalkan margin adalah kerugian engsel.

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \quad (1)$$

Pada persamaan 1 biaya adalah 0 jika nilai prediksi dan nilai aktual memiliki tanda yang sama. Jika tidak, kami kemudian menghitung nilai kerugiannya. Kami juga menambahkan parameter regularisasi fungsi biaya. Tujuan dari parameter regularisasi adalah untuk menyeimbangkan maksimalisasi margin dan kerugian. Setelah menambahkan parameter regularisasi, fungsi biaya terlihat pada persamaan 2.

$$\min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)_+ \quad (2)$$

Sekarang kita memiliki fungsi kerugian, kita mengambil turunan parsial sehubungan dengan bobot untuk menemukan gradien. Menggunakan gradien, kita dapat memperbarui bobot kita seperti pada persamaan 3.

$$\frac{\delta}{\delta w_k} \lambda \|w\|^2 = 2\lambda w_k$$

$$\frac{\delta}{\delta w_k} (1 - y_i \langle x_i, w \rangle)_+ = \begin{cases} 0, & \text{if } y_i \langle x_i, w \rangle \geq 1 \\ -y_i x_{ik}, & \text{else} \end{cases} \quad (3)$$

Pada persamaan 4, ketika tidak ada kesalahan klasifikasi, yaitu model kami memprediksi dengan benar kelas titik data kami, kami hanya perlu memperbarui gradien dari parameter regularisasi.

$$w = w - \alpha \cdot (2\lambda w) \quad (4)$$

Pada persamaan 5, ketika ada kesalahan klasifikasi, yaitu model kami membuat kesalahan pada prediksi kelas titik data kami,

kami menyertakan kerugian bersama dengan parameter regularisasi untuk melakukan pembaruan gradien.

$$w = w + \alpha \cdot (y_i \cdot x_i - 2\lambda w) \quad (5)$$

#### 4. HASIL DAN ANALISIS

Data yang telah diperoleh, diolah menggunakan software RapidMiner seperti pada Gambar 1. Pada tahap *cross validation*, data diolah dengan metode *support vector machine*.



Gambar 1. Proses prediksi

Pada Gambar 2, dilakukan prediksi pada data ke-492 dengan proses yang telah terbentuk pada RapidMiner. Hasil menunjukkan, data ke-492 termasuk pada kelas “positive”.

Case	predicted	confidence	actual	correct	error	weight	status
1	positive	0.853	positive	1	0	0	1

Gambar 2. Contoh prediksi pada data 492

Pada hasil evaluasi analisis sentimen dengan menggunakan confusion matrix, diperoleh hasil akurasi sebesar 63.5% seperti yang ditunjukkan pada Gambar 3. Hasil true negative ditemukan sebesar 26, false negative sebesar 4, false positive sebesar 69, dan true positive ditemukan sebesar 101. Dari Gambar 2, diperoleh hasil sensitivity sebesar 96,19%, specificity sebesar 27,37%, negative predictive value sebesar 86,67, dan precision sebesar 59,41%.

accuracy: 63.50% +/- 6.26% (micro average: 63.50%)

	true negative	true positive	class precision
pred. negative	26	4	86.67%
pred. positive	69	101	59.41%
class recall	27.37%	96.19%	

Gambar 3. Hasil evaluasi *confusion matrix*

#### 5. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

Algoritma *support vector machine* dapat digunakan untuk mengklasifikasikan *tweet* kedalam positif atau negatif terutama *tweet* mengenai pasca bencana

Berdasarkan hasil dan pembahasan, akurasi yang didapatkan dari algoritma *support vector machine* adalah sebesar 63.5%.

Kemudian penelitian ini berguna untuk penelitian selanjutnya, peneliti dapat menggunakan metode lainnya sehingga dapat dilakukan perbandingan dari masing-masing metode. Lexicon analisis sentimen yang berbeda juga dapat dilakukan sehingga analisis lebih lanjut dapat diperoleh.

#### 6. DAFTAR PUSTAKA

- A. S. Ahmad Fathan Hidayatullah, "Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik pada Twitter," in Seminar Nasional Informatika , Yogyakarta, 2014.
- Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation. CITEE 2014 , 39-43.
- Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. IEEE Intelligent Systems , 25, 46–53.
- Mohammad, S. M., Kiritchenko, S., & Zhu, X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. Procee-dings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013). Atlanta, USA. 2013.
- Pirna, Mironela. (2018). Sentiment analysis for the tweets that contain the word "earthquake". ECAI 2018 - International Conference – 10th Edition.
- Pradhan, Sameer S., et al. "Shallow semantic

- parsing using support vector machines." Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004. 2004.
- S. W. S. A. D. Faishol Nur Huda, "Analisis Sentimen masyarakat Terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," ITSMART, vol. 2, no. 2, pp. 35 - 42, 2013.
- M. A. Assuja and Saniati, "Analisis Sentimen Tweet Menggunakan," J. Tek., vol. 10, no. 2, pp. 23–28, 2016.
- A. P. Wijaya, H. A. Santoso, J. T. Informatika, U. Dian, and N. Semarang, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," J. Appl. Intell. Syst., vol. 1, no. 1, pp. 48–55, 2016.
- Vo, Bao-Khanh H., Collier, Nigel. (2013). Twitter Emotion Analysis in Earthquake Situations. IJCLA Vol. 4, No. 1, Jan-Jun 2013, Pp. 159–173.
- I. D. I, I. Amirulloh, and D. Rosiyadi, "Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda Naive Bayes," J. Electr. Electron. Eng., vol. 1, no. 2, pp. 38–45, 2016.
- D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter," J. Tekno Kompak, vol. 14, no. 2, pp. 86–91, 2020.