Technologica

Vol.5(2): Hal 250-263 (Juli 2026)

ISSN: 2827-9492 (*Online*)

Model Klasifikasi *Tweet* Bencana Alam Gunung Meletus Menggunakan Algoritma *Long Short-Term Memory*

Classification Model of Volcanic Disaster Tweets using the Long Short-Term Memory Algorithm

Satria Nugraha Saputra[1], Febrian Wahyu Christanto[*,2], Eryan Ahmad Firdaus[3], Aldian Umbu Tamu Ama[4]

[1,2]Program Studi S1 Teknik Informatika, Universitas Semarang, Jawa Tengah, Indonesia
[3]Program Studi S1 Informatika, Universitas Pertahanan, Jawa Barat, Indonesia
[4]Program Studi S1 Rekayasa Perangkat Lunak, Universitas Pignatelli Triputra, Jawa Tengah, Indonesia

[*]Penulis Korespondensi
Email: febrian.wahyu.christanto@usm.ac.id

Abstrak. Media sosial seperti Twitter (X) menjadi salah satu sumber informasi *real-time* yang banyak digunakan masyarakat untuk menyampaikan kondisi dan perkembangan situasi saat terjadi bencana alam. Namun data media sosial memiliki karakteristik tidak terstruktur, mengandung *noise*, dan menggunakan bahasa yang dinamis sehingga diperlukan pendekatan klasifikasi teks otomatis untuk mengidentifikasi informasi kebencanaan secara efektif. Tujuan penelitian ini adalah mengembangkan model klasifikasi *tweet* mengenai bencana alam gunung meletus menggunakan algoritma *Long Short-Term Memory* (LSTM) melalui tahapan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM). *Dataset* yang digunakan berasal dari "*Disaster Tweets Dataset*" pada *platform* Kaggle dengan total 11.370 data *tweet* yang terdiri dari kategori bencana dan non-bencana. Tahapan penelitian meliputi data *preprocessing*, *tokenization*, *text representation* menggunakan *embedding layer*, pemodelan menggunakan LSTM, serta evaluasi model menggunakan *confusion matrix*. Berdasarkan hasil pengujian, model memperoleh nilai *accuracy* sebesar 89%, *precision* sebesar 89%, *recall* sebesar 39%, dan *F1-score* sebesar 54% pada pembagian data *training* dan *testing* dengan rasio 80:20. Performa model yang ditunjukkan oleh nilai *accuracy* dan *precision* yang tinggi mengindikasikan kemampuan yang baik dalam mengklasifikasikan *tweet* bencana. Meskipun demikian, nilai *recall* yang masih rendah mengisyaratkan bahwa model belum mampu mendeteksi seluruh *tweet* bencana yang terdapat dalam *dataset* secara optimal. Penelitian ini menunjukkan bahwa pemanfaatan algoritma LSTM dan data media sosial memiliki potensi untuk mendukung pengembangan sistem *monitoring* informasi kebencanaan secara otomatis dan *real-time*.
Kata kunci: Bencana Alam Gunung Meletus, *Long Short-Term Memory* (LSTM), *Natural Language Processing* (NLP), Twitter (X), CRISP-DM

Abstract. *Social media platforms like Twitter (X) serve as widely used sources of real-time information for the public to report conditions and developments during natural disasters. However, social media data is characterized by its unstructured nature, noise, and dynamic language use; thus, automated text classification approaches are required to effectively identify disaster-related information. This research aims to develop a classification model for tweets regarding volcanic eruptions using the Long Short-Term Memory (LSTM) algorithm, following the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The dataset*

utilized is the "Disaster Tweets Dataset" from Kaggle, comprising 11,370 tweets categorized into disaster and non-disaster classes. The research process encompasses data preprocessing, tokenization, text representation via an embedding layer, LSTM-based modeling, and model evaluation using a confusion matrix. Test results show that the model achieved an accuracy of 89%, precision of 89%, recall of 39%, and an F1-score of 54% using an 80:20 training-to-testing data split. The model's performance—demonstrated by high accuracy and precision—indicates a strong capability in classifying disaster-related tweets. Nevertheless, the low recall score suggests that the model has not yet optimally detected all disaster-related tweets within the dataset. This research demonstrates that leveraging the LSTM algorithm and social media data holds potential for supporting the development of automated, real-time disaster information monitoring systems.
Keywords: *Volcanic Eruption Disaster, Long Short-Term Memory (LSTM), Natural Language Processing (NLP), Twitter (X), CRISP-DM*

1. Pendahuluan

Posisi geografis serta karakteristik geologis Indonesia yang terletak di kawasan Cincin Api Pasifik (*Ring of Fire*) menyebabkan negara ini memiliki tingkat risiko yang tinggi terhadap berbagai jenis bencana alam (Christanto *et al*., 2025). Sebagai negara yang memiliki tingkat kerawanan bencana yang tinggi, Indonesia kerap mengalami berbagai bencana alam termasuk gempa bumi, banjir, tanah longsor, dan erupsi gunung api. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) tercatat sebanyak 3.946 kejadian bencana sepanjang tahun 2024 yang menimbulkan dampak yang signifikan terhadap aspek sosial, ekonomi, dan lingkungan (BNPB, 2025). Kondisi tersebut menunjukkan pentingnya sistem informasi kebencanaan yang mampu mendukung proses mitigasi, tanggap darurat, dan penyebaran informasi secara cepat dan akurat.

Pesatnya perkembangan teknologi komunikasi dan media sosial telah mentransformasi mekanisme penyebaran informasi kebencanaan. Dalam konteks ini masyarakat tidak lagi hanya berfungsi sebagai penerima informasi tetapi juga berkontribusi sebagai penghasil dan penyebar informasi melalui *platform* media sosial seperti Twitter (X) (Fauzianto & Supatman, 2023). Informasi yang dibagikan pengguna media sosial umumnya mencakup lokasi kejadian, kondisi lapangan, kebutuhan korban, hingga perkembangan situasi pasca bencana. Karakteristik media sosial yang bersifat *real-time* memungkinkan informasi terkait bencana diperoleh lebih cepat dibandingkan laporan resmi dari lembaga pemerintah (Pandunata *et al*., 2022). Oleh karena itu data media sosial memiliki potensi besar untuk dimanfaatkan sebagai sumber informasi pendukung dalam mitigasi bencana dan peningkatan *situational awareness*.

Meskipun demikian data yang berasal dari media sosial memiliki berbagai tantangan seperti struktur data yang tidak baku, penggunaan bahasa informal, singkatan, *noise*, serta potensi penyebaran informasi yang tidak relevan. Kondisi tersebut menyebabkan proses identifikasi informasi kebencanaan secara manual menjadi kurang efektif terutama ketika volume data meningkat secara signifikan pada saat terjadi bencana (Sulthan *et al*., 2021). Dengan demikian, diperlukan pendekatan analisis teks berbasis otomatisasi yang mampu melakukan klasifikasi informasi pada data media sosial secara cepat, efisien, dan konsisten sehingga dapat mendukung proses identifikasi informasi kebencanaan secara lebih efektif.

Natural Language Processing (NLP) merupakan salah satu pendekatan dalam bidang kecerdasan buatan yang digunakan untuk mengolah data teks. Dengan memanfaatkan NLP, komputer dapat memahami, memproses, dan menganalisis bahasa alami sehingga informasi yang

relevan dari data media sosial dapat diidentifikasi dan diekstraksi secara otomatis (Nofiyanti & Haryanto, 2021). Dalam penelitian kebencanaan, NLP telah banyak dimanfaatkan untuk analisis sentimen, klasifikasi informasi bencana, serta identifikasi kebutuhan masyarakat terdampak (Husada & Toba, 2020). Selain daripada NLP, penerapan teknik *deep learning* seperti *Long Short-Term Memory* (LSTM) dinilai mampu meningkatkan kemampuan model dalam memahami konteks kalimat dan hubungan antar kata pada data teks dibandingkan metode klasifikasi konvensional (Minaee *et al*., 2021), (Kuz *et al*., 2025).

Beberapa penelitian sebelumnya telah menerapkan metode klasifikasi untuk analisis data media sosial terkait kebencanaan. Salah satu penelitian menggunakan metode *K-Nearest Neighbor* (K-NN) dan *Decision Tree* untuk analisis sentimen pasca bencana alam dengan tingkat *accuracy* sebesar 79%. Namun penelitian tersebut masih menunjukkan nilai *recall* yang relatif rendah sehingga model belum optimal dalam mendeteksi seluruh data positif (Shahputri & Yamasari, 2023). Penelitian lain menggunakan metode *Naive Bayes Classifier* untuk analisis potensi bencana tanah longsor dengan hasil *accuracy* sebesar 82,61%. Meskipun hasil *accuracy* cukup baik, model masih memiliki keterbatasan dalam menangani kompleksitas konteks bahasa pada data teks media social (Wahyuni & Susanti, 2023). Analisis dari penelitian sebelumnya masih terdapat kebutuhan untuk mengembangkan model klasifikasi yang mampu meningkatkan kemampuan identifikasi informasi kebencanaan secara lebih efektif.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan model klasifikasi *tweet* bencana alam dengan memanfaatkan algoritma *Long Short-Term Memory* (LSTM) yang diimplementasikan menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Kerangka kerja CRISP-DM diterapkan untuk memastikan setiap tahapan penelitian dilaksanakan secara sistematis mulai dari pemahaman terhadap permasalahan, persiapan dan pengolahan data, pembangunan model, hingga evaluasi kinerja model. Penelitian ini bertujuan untuk menganalisis serta mengklasifikasikan *tweet* yang berkaitan dengan bencana alam gunung meletus sehingga informasi penting yang tersebar di media sosial dapat diidentifikasi secara otomatis. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem informasi kebencanaan berbasis media sosial yang lebih cepat, akurat, dan efektif guna mendukung upaya mitigasi serta penanganan tanggap darurat bencana.

2. Metode

2.1 *Dataset* Penelitian

Pada penelitian ini digunakan dataset publik "*Disaster Tweets*" yang diunduh dari *platform* Kaggle sebagai sumber data penelitian. Gambaran mengenai *dataset* yang digunakan ditampilkan pada Gambar 1.

Gambar 1. *Dataset Disaster Tweet* (Stepanenko, 2020)

Dataset ini dipilih karena menyediakan *data tweet* berlabel yang berkaitan dengan bencana (*disaster*) dan non-bencana (*non-disaster*) sehingga sesuai untuk penelitian klasifikasi teks berbasis *Natural Language Processing* (NLP) (Airlangga, 2024). Selain itu *dataset* ini telah banyak digunakan sebagai *benchmark* dalam penelitian klasifikasi teks pada media social (Qutaishat & Li, 2026). Dataset terdiri dari 11.370 *data tweet* dengan beberapa atribut utama yaitu *id*, *keyword*, *location*, *text*, dan *target*. Variabel target digunakan sebagai label klasifikasi dimana nilai 1 menunjukkan *tweet* berkaitan dengan bencana sedangkan nilai 0 menunjukkan *tweet* tidak berkaitan dengan bencana. Distribusi kelas pada *dataset* ditunjukkan pada Tabel 1.

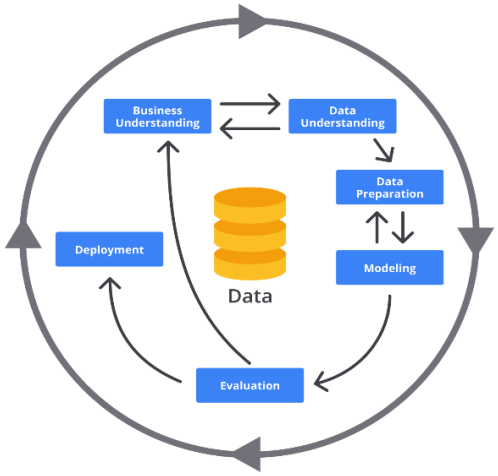
Tabel 1. Detail *Dataset Disaster Tweet*

Kelas	**Jumlah *Tweet***
Disaster (1)	9.096
Non-Disaster (0)	2.274
Total	11.370

Berdasarkan distribusi tersebut *dataset* memiliki karakteristik *imbalanced dataset* karena jumlah *tweet* kelas positif lebih dominan dibandingkan kelas negatif. Ketidakseimbangan data ini berpotensi memengaruhi performa model terutama pada nilai *recall* dan kemampuan model dalam mengenali seluruh *tweet* yang berkaitan dengan bencana.

2.2 Metode Penelitian

Sebagai kerangka kerja penelitian, riset ini mengadopsi metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Pemilihan metodologi tersebut didasarkan pada kemampuannya dalam menyediakan alur kerja yang sistematis dan terstruktur untuk proses pengolahan data hingga pengembangan model *machine learning*. CRISP-DM mencakup enam tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Rangkaian tahapan penelitian berdasarkan metodologi CRISP-DM ditunjukkan pada Gambar 2.



Gambar 2. Metodologi CRISP-DM (Martinez-Plumed *et al*., 2021)

Tahap *Business Understanding* difokuskan pada identifikasi permasalahan dan analisis kebutuhan sistem klasifikasi informasi kebencanaan yang bersumber dari media sosial. Pada tahap *Data Understanding* dilakukan analisis terhadap karakteristik *dataset*, distribusi data, serta kualitas data sebagai dasar dalam proses pemodelan. Selanjutnya, tahap *Data Preparation* meliputi pembersihan, transformasi, dan persiapan data teks sehingga dapat digunakan secara optimal dalam pembentukan model. Proses pembangunan model dilakukan pada tahap *Modeling* dengan

memanfaatkan algoritma *Long Short-Term Memory* (LSTM). Kinerja model kemudian dievaluasi pada tahap *Evaluation* menggunakan *confusion matrix* dan berbagai metrik evaluasi. Tahap *Deployment* merupakan tahap akhir yang bertujuan mengimplementasikan model ke dalam *dashboard* pemantauan informasi kebencanaan.

2.3 Data Preparation

Pada tahap *Data Preparation* ini dilakukan proses pembersihan dan standarisasi *data tweet* sebelum digunakan pada proses pelatihan model (Hidayati *et al*., 2021). Tahap ini penting dilakukan karena data yang berasal dari media sosial umumnya mengandung berbagai *noise* atau informasi tidak relevan yang dapat memengaruhi kualitas hasil klasifikasi. Selain itu *data tweet* memiliki karakteristik bahasa yang tidak terstruktur dan sering menggunakan singkatan maupun variasi penulisan sehingga diperlukan proses *preprocessing* agar data menjadi lebih konsisten dan mudah diproses oleh model *machine learning*. Proses *preprocessing* diawali dengan tahapan *case folding* yaitu mengonversi seluruh karakter teks menjadi huruf kecil (*lowercase*) guna menyeragamkan representasi kata dalam *dataset*. Tahapan ini bertujuan untuk menghindari perbedaan representasi kata akibat variasi penggunaan huruf kapital. Selanjutnya, dilakukan proses *cleaning* dengan menghapus berbagai elemen yang tidak relevan terhadap proses klasifikasi, seperti URL, *mention* (*@username*), *hashtag* (#), emoji, angka, tanda baca, serta karakter khusus lainnya, sehingga kualitas data menjadi lebih baik untuk tahap pemodelan. Tahap ini dilakukan bertujuan untuk mengurangi *noise* pada data sehingga informasi utama dalam *tweet* dapat dipertahankan.

Setelah tahap *cleaning* dilakukan proses *tokenization* untuk memecah teks menjadi *token* individual sebagai dasar pembentukan representasi teks. Tahap berikutnya adalah *stopword removal* yaitu proses menghilangkan kata-kata umum yang memiliki bobot informasi rendah dan tidak berpengaruh secara signifikan terhadap proses klasifikasi, seperti "dan", "yang", "di", serta "ke". Penghapusan *stopword* bertujuan untuk meningkatkan kualitas fitur dengan mempertahankan kata-kata yang lebih representatif terhadap konteks kebencanaan. Selanjutnya, diterapkan proses *stemming* untuk mengonversi setiap kata ke bentuk dasarnya sehingga variasi kata yang memiliki makna serupa dapat diseragamkan. Sebagai contoh kata “berjalan”, “jalanan”, dan “berjalanlah” dapat direduksi menjadi kata dasar “jalan”. Proses *stemming* membantu meningkatkan konsistensi representasi kata dalam data teks sehingga dapat meningkatkan efektivitas proses klasifikasi. Selain itu dilakukan pula proses *duplicate removal* menggunakan metode *drop_duplicates*( ) untuk menghapus *data tweet* yang identik. Tahap ini bertujuan mengurangi redundansi data yang dapat memengaruhi performa model selama proses pelatihan. Setelah seluruh tahapan *preprocessing* selesai dilakukan maka data hasil *preprocessing* kemudian digunakan sebagai *input* pada tahap pemodelan menggunakan algoritma *Long Short-Term Memory* (LSTM) (Dhewayani *et al*., 2022), (Mubarak *et al*., 2024). Dalam Tabel 2 adalah data hasil *preprocessing* terhadap *data tweet*.

Tabel 2. Hasil Tahapan *Preprocessing*

Tahap	Hasil
Raw Tweet	*“RT @news: Volcano eruption happened!!! #disaster”*
Cleaning	*“volcano eruption happened disaster”*
Tokenization	*[volcano, eruption, happened, disaster]*
Stopword Removal	*[volcano, eruption, happened, disaster]*
Stemming	*[volcano, erupt, happen, disaster]*

Proses *preprocessing* bertujuan untuk meningkatkan kualitas data teks melalui serangkaian tahapan pengolahan sehingga model dapat mengidentifikasi dan mempelajari pola informasi secara lebih optimal. Tahapan *preprocessing* yang baik dapat membantu meningkatkan kualitas representasi fitur pada data teks sehingga proses pembelajaran model menjadi lebih optimal. Selain meningkatkan performa klasifikasi, *preprocessing* juga membantu mengurangi kompleksitas data dan risiko kesalahan prediksi akibat keberadaan elemen teks yang tidak relevan.

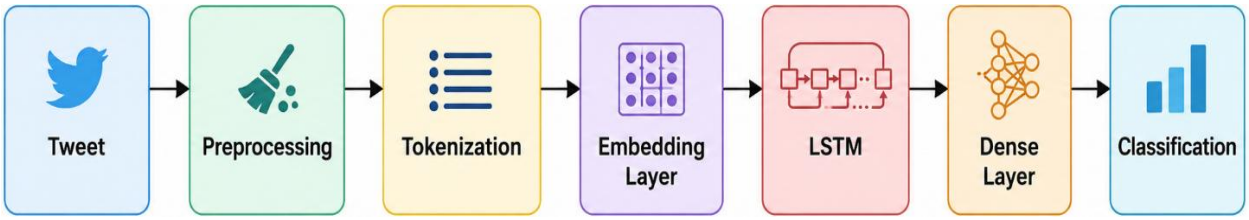
2.4 Pemodelan Menggunakan *Long Short-Term Memory* (LSTM)

Dalam penelitian ini algoritma *Long Short-Term Memory* (LSTM) digunakan sebagai model klasifikasi *tweet* bencana alam. LSTM merupakan salah satu varian *Recurrent Neural Network* (RNN) yang dikembangkan untuk mengatasi permasalahan *long-term dependency* pada data sekuensial khususnya data berbasis teks. Dibandingkan dengan RNN konvensional, LSTM memanfaatkan mekanisme *memory cell* serta *input gate*, *forget gate*, dan *output gate* untuk mempertahankan atau mengendalikan aliran informasi selama proses pembelajaran. Mekanisme tersebut memungkinkan model menangkap hubungan kontekstual antarkata secara lebih efektif, sehingga mampu meningkatkan kinerja klasifikasi pada data *tweet* yang umumnya bersifat singkat, tidak terstruktur, menggunakan bahasa tidak formal, serta memiliki variasi penulisan yang tinggi.

Arsitektur model yang diusulkan pada penelitian ini dibangun melalui beberapa lapisan utama meliputi *Embedding Layer*, *Spatial Dropout*, LSTM *Layer*, *Dense Layer*, *Dropout Layer*, dan *Output Layer*. Setiap lapisan memiliki fungsi yang berbeda dalam proses representasi fitur, pembelajaran pola sekuensial, regularisasi model, hingga menghasilkan keluaran berupa hasil klasifikasi (Fabillah *et al*., 2023), (Ashari & Suhendar, 2024). *Embedding layer* bertugas mentransformasikan setiap kata ke dalam ruang vektor berdimensi tertentu sehingga representasi semantik antar kata dapat dipelajari oleh model. Representasi vektor ini memungkinkan kata-kata yang memiliki makna atau konteks serupa dipetakan ke posisi yang berdekatan dalam ruang fitur, sehingga meningkatkan kemampuan model dalam memahami konteks kalimat. Representasi vektor tersebut membantu model memahami kemiripan makna antar kata dalam data teks. Setelah proses *embedding* maka akan diterapkan *Spatial Dropout* untuk mengurangi risiko *overfitting* pada representasi fitur teks dengan cara menghilangkan sebagian *neuron* secara acak selama proses pelatihan. *Layer* utama dalam model ini adalah LSTM dengan 200 unit yang digunakan untuk menangkap dependensi konteks jangka panjang pada urutan kata dalam *tweet*. Penggunaan jumlah unit tersebut bertujuan agar model mampu mempelajari pola bahasa dan hubungan antar kata secara lebih optimal. Selanjutnya hasil keluaran dari layer LSTM diteruskan ke *Dense Layer* untuk proses klasifikasi. Pada tahap ini digunakan pula *Dropout Layer* untuk meningkatkan kemampuan generalisasi model dan mengurangi kemungkinan model mengalami *overfitting* selama proses pelatihan. *Layer* terakhir yaitu *Output Layer* digunakan untuk menghasilkan prediksi klasifikasi *tweet* ke dalam kategori bencana (*disaster*) atau non-bencana (*non-disaster*).

Untuk membangun dan mengevaluasi model, *dataset* dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan rasio 80:20. Sebanyak 80% data dimanfaatkan pada tahap pelatihan untuk mempelajari pola dalam *dataset* sedangkan 20% sisanya digunakan pada tahap pengujian untuk mengukur kemampuan generalisasi dan kinerja model terhadap data yang belum pernah dipelajari sebelumnya (Hidayat *et al*., 2026). Pemilihan rasio tersebut dilakukan karena merupakan konfigurasi yang umum digunakan dalam penelitian klasifikasi teks dan dinilai mampu memberikan keseimbangan antara jumlah data pelatihan dan data evaluasi. Model dilatih

selama 20 *epoch* untuk memperoleh performa klasifikasi yang optimal. Alur proses pemodelan penelitian ditunjukkan pada Gambar 3 berikut.



Gambar 3. Alur Proses Pemodelan Klasifikasi *Tweet* Menggunakan Algoritma LSTM

Gambar 3 menunjukkan alur proses pemodelan klasifikasi *tweet* bencana alam menggunakan algoritma *Long Short-Term Memory* (LSTM). Proses dimulai dari pengambilan *data tweet* yang kemudian melalui tahap *preprocessing* untuk membersihkan data dari *noise*. Setelah itu dilakukan *tokenization* untuk memecah teks menjadi *token* kata sebelum direpresentasikan ke dalam bentuk vektor numerik melalui *embedding layer*. Selanjutnya data diproses menggunakan model LSTM untuk mempelajari hubungan konteks antar kata pada *tweet*. Hasil keluaran dari LSTM diteruskan ke *dense layer* untuk melakukan proses klasifikasi akhir sehingga *tweet* dapat dikategorikan ke dalam kelas bencana (*disaster*) atau non-bencana (*non-disaster*).

2.5 Evaluasi Model

Tahap evaluasi bertujuan untuk mengukur kinerja model klasifikasi yang telah dibangun melalui analisis *confusion matrix*. Pendekatan ini digunakan untuk menilai kemampuan model dalam mengidentifikasi dan mengklasifikasikan *tweet* ke dalam kelas bencana (*disaster*) maupun non-bencana (*non-disaster*) secara tepat. Agar evaluasi kinerja model dilakukan secara lebih komprehensif dalam penelitian ini menerapkan beberapa metrik evaluasi yaitu *accuracy*, *precision*, *recall*, dan *F1-score* yang masing-masing memberikan informasi mengenai tingkat ketepatan klasifikasi, presisi prediksi, sensitivitas model, serta keseimbangan antara *precision* dan *recall* (Opitz, 2024). *Accuracy* merupakan metrik evaluasi yang digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data pada proses pengujian. Metrik ini memberikan gambaran mengenai kemampuan model dalam melakukan klasifikasi secara keseluruhan, sehingga semakin tinggi nilai *accuracy* maka semakin baik tingkat ketepatan prediksi yang dihasilkan oleh model.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (1)$$

Precision merupakan metrik evaluasi yang digunakan untuk mengukur tingkat ketepatan prediksi model terhadap kelas positif. Metrik ini menunjukkan proporsi prediksi positif yang benar yaitu *tweet* yang diprediksi sebagai bencana dan benar-benar termasuk dalam kategori bencana. Nilai *precision* yang tinggi mengindikasikan bahwa model mampu meminimalkan terjadinya *false positive* sehingga prediksi yang dihasilkan menjadi lebih andal.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan metrik evaluasi yang digunakan untuk mengukur sensitivitas model dalam mendeteksi data pada kelas positif. Metrik ini menunjukkan proporsi *tweet* bencana yang berhasil diidentifikasi secara benar dibandingkan dengan seluruh *tweet* yang benar-benar termasuk dalam kategori bencana pada *dataset*. Nilai *recall* yang tinggi mengindikasikan bahwa model mampu

meminimalkan terjadinya *false negative* sehingga semakin sedikit *tweet* bencana yang gagal terdeteksi.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Selain itu digunakan pula nilai *F1-score* untuk mengukur keseimbangan antara *precision* dan *recall*. Metrik ini penting digunakan terutama pada *dataset* yang tidak seimbang (*imbalanced dataset*) karena mampu memberikan evaluasi yang lebih representatif terhadap performa model klasifikasi.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Confusion matrix merupakan metode evaluasi yang menyajikan hasil klasifikasi ke dalam empat komponen utama yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Nilai TP menunjukkan banyaknya data pada kelas positif yang berhasil diidentifikasi dengan benar oleh model sedangkan TN menunjukkan banyaknya data pada kelas negatif yang juga berhasil diklasifikasikan secara benar. Sebaliknya, FP merepresentasikan jumlah data pada kelas negatif yang keliru diklasifikasikan sebagai kelas positif sementara FN menunjukkan jumlah data pada kelas positif yang salah diklasifikasikan sebagai kelas negatif. (Rainio *et al*., 2024).

Penerapan beberapa metrik evaluasi bertujuan untuk memperoleh penilaian yang lebih komprehensif terhadap kinerja model khususnya pada permasalahan klasifikasi *tweet* bencana dengan distribusi kelas yang tidak seimbang (*class imbalance*). Dalam kondisi tersebut metrik *accuracy* saja belum memadai sebagai indikator performa model karena nilai *accuracy* yang tinggi tidak selalu mencerminkan kemampuan model dalam mengidentifikasi seluruh data pada kelas positif. Oleh karena itu, metrik *precision*, *recall*, dan *F1-score* turut digunakan untuk melengkapi proses evaluasi sehingga kemampuan model dalam melakukan klasifikasi secara akurat, konsisten, dan seimbang terhadap kedua kelas dapat dinilai secara lebih objektif.

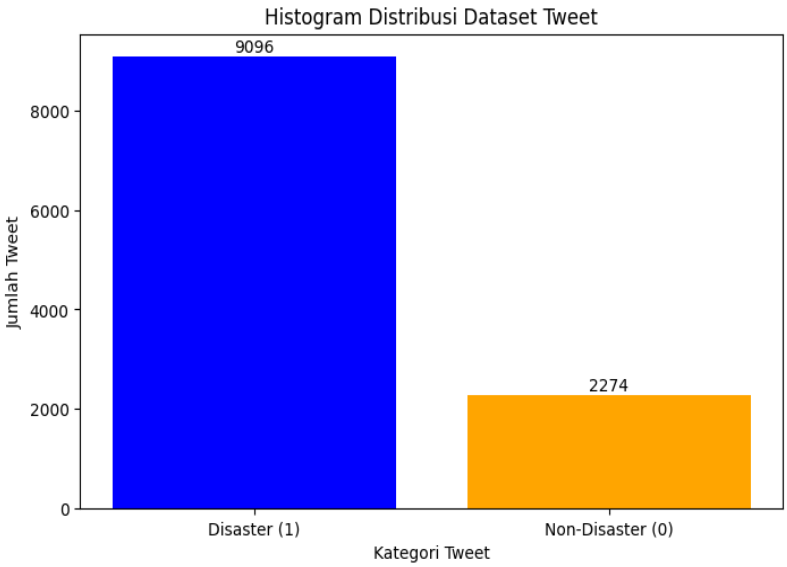
3. Hasil dan Pembahasan

Penelitian ini menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dalam proses klasifikasi *tweet* bencana alam menggunakan algoritma *Long Short-Term Memory* (LSTM). Tahapan penelitian dimulai dari proses *Business Understanding* hingga *Deployment* untuk menghasilkan model klasifikasi *tweet* yang mampu mengidentifikasi informasi terkait bencana alam gunung meletus dari media sosial Twitter (X).

3.1 Data Understanding

Tahap *Data Understanding* bertujuan untuk mengidentifikasi dan menganalisis karakteristik *dataset* yang digunakan sebagai dasar dalam proses pemodelan. Hasil analisis karakteristik *dataset* disajikan pada Tabel 1. Berdasarkan distribusi *dataset* tersebut terlihat bahwa jumlah data *tweet* kategori bencana lebih dominan dibandingkan kategori non-bencana. Kondisi ini menunjukkan bahwa *dataset* memiliki karakteristik *imbalanced dataset* yang dapat memengaruhi performa model terutama dalam mendeteksi seluruh data positif.

Selanjutnya dilakukan visualisasi distribusi data menggunakan histogram untuk mengetahui persebaran kelas pada *dataset*. Histogram hasil distribusi data ditunjukkan pada Gambar 4.

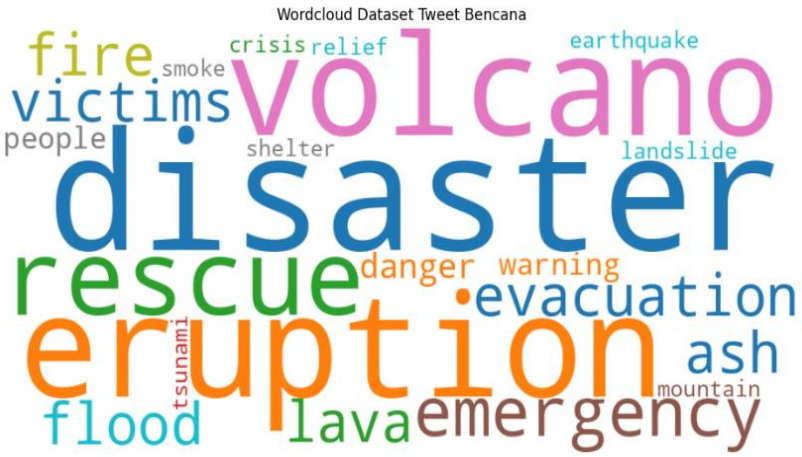


Gambar 4. Histogram Distribusi *Dataset Tweet*

Gambar 4 menunjukkan bahwa jumlah *tweet* kategori *disaster* jauh lebih banyak dibandingkan kategori *non-disaster*. Ketidakseimbangan distribusi data tersebut dapat menyebabkan model cenderung lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas.

3.2 Data Preparation

Pada tahap *Data Preparation* dilakukan serangkaian proses *preprocessing* yang terdiri atas *case folding, cleaning, tokenization, stopword removal, stemming,* dan *duplicate removal*. Hasil dari setiap tahapan *preprocessing* disajikan pada Tabel 2. Proses tersebut meningkatkan kualitas dan konsistensi data teks dengan menghilangkan elemen yang tidak relevan serta menyeragamkan representasi kata sehingga data menjadi lebih representatif dan siap digunakan dalam proses pelatihan model. Selain histogram, *data tweet* juga divisualisasikan menggunakan *wordcloud* untuk mengetahui kata yang paling sering muncul pada *dataset*. Visualisasi *wordcloud* ditunjukkan pada Gambar 5.



Gambar 5. *Wordcloud Dataset Tweet* Bencana

Berdasarkan hasil visualisasi *wordcloud* terlihat bahwa kata-kata seperti "*fire*", "*eruption*", "*evacuation*", "*disaster*", dan "*people*" memiliki frekuensi kemunculan yang tinggi pada *tweet* kategori bencana. Hal tersebut menunjukkan bahwa *dataset* mengandung informasi yang relevan terhadap kondisi kebencanaan dan dapat digunakan sebagai dasar dalam proses klasifikasi teks.

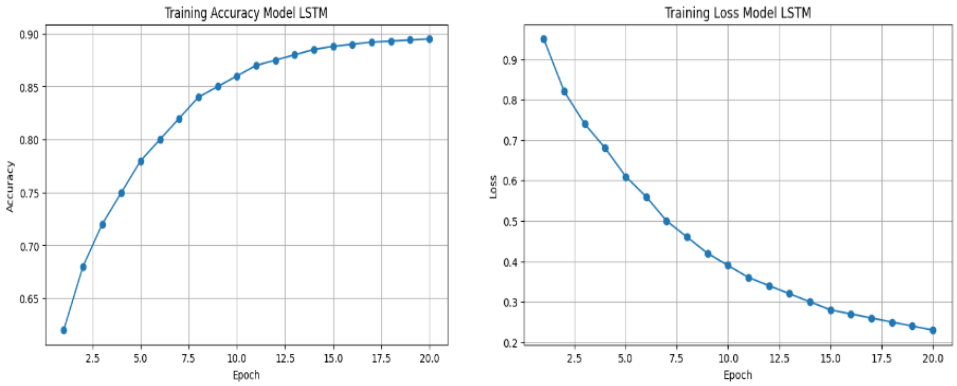
3.3 Pemodelan Menggunakan LSTM

Dari alur proses pemodelan klasifikasi *tweet* menggunakan algoritma LSTM yang telah ditentukan pada Gambar 3 didapatkan rincian arsitektur model LSTM yang digunakan ditunjukkan pada Tabel 3 berikut.

Tabel 3. Arsitektur Model LSTM

Layer (Type)	***Output Shape***	***Parameter***
Embedding	*(None, None, 2)*	10.000
SpatialDropout1D	*(None, None, 2)*	0
LSTM	*(None, 200)*	162.400
Dropout	*(None, 200)*	0
Dense	*(None, 150)*	30.150
Dropout	*(None, 150)*	0
Dense	*(None, 100)*	15.100
Dropout	*(None, 100)*	0
Dense	*(None, 1)*	101
Total Parameter	-	217.751
Embedding	*(None, None, 2)*	10.000

Berdasarkan arsitektur model pada Tabel 3, model LSTM memiliki total 217.751 parameter yang seluruhnya termasuk ke dalam kategori *trainable parameters*. Jumlah parameter tersebut mencerminkan kapasitas model yang cukup untuk mempelajari representasi data teks serta hubungan kontekstual antarkata dalam *tweet*. Penggunaan beberapa *dense layer* juga berperan dalam meningkatkan kemampuan model untuk mengekstraksi representasi fitur yang lebih kompleks sebelum menghasilkan keputusan klasifikasi. Pada tahap pelatihan, *dataset* dibagi menjadi *training set* dan *testing set* dengan rasio 80:20. Sebanyak 80% data digunakan untuk membangun model sedangkan 20% sisanya dimanfaatkan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dipelajari sebelumnya. Pemilihan rasio tersebut didasarkan pada pertimbangan untuk memperoleh keseimbangan antara proses pembelajaran dan evaluasi sehingga model dapat dilatih secara optimal sekaligus menghasilkan penilaian performa yang lebih representatif. Model dilatih selama 20 *epoch* untuk memperoleh performa klasifikasi yang optimal dan meminimalkan kesalahan prediksi selama proses pelatihan. Selain itu proses pelatihan dilakukan secara bertahap agar model mampu menyesuaikan bobot *parameter* secara lebih stabil pada setiap iterasi pembelajaran. Untuk Grafik *accuracy* dan *loss* selama proses pelatihan model ditunjukkan pada Gambar 6.



Gambar 6. *Wordcloud Dataset Tweet* Bencana

Berdasarkan hasil yang ditampilkan pada Gambar 6 nilai *accuracy* menunjukkan tren peningkatan sedangkan nilai *loss* mengalami penurunan sepanjang proses pelatihan. Pola tersebut

mengindikasikan bahwa model secara bertahap mampu mengoptimalkan proses pembelajaran dalam mengenali pola pada data sehingga kinerja model meningkat selama tahap *training*.

3.4 Evaluasi Model

Tahap evaluasi bertujuan untuk menilai kinerja model klasifikasi melalui analisis *confusion matrix*. Penilaian performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* sehingga diperoleh gambaran yang komprehensif mengenai kemampuan model dalam melakukan klasifikasi. Hasil evaluasi tersebut disajikan pada Tabel 4.

Tabel 4. Hasil *Confusion Matrix*

Prediksi Aktual	***Negative* (0)**	***Positive* (1)**
Negative (0)	1860 (TN)	18 (FP)
Positive (1)	243 (FN)	153 (TP)

Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model menghasilkan nilai *True Negative* (TN) sebesar 1860 dan *True Positive* (TP) sebesar 153 yang mengindikasikan bahwa sebagian besar data negatif serta sebagian data positif berhasil diklasifikasikan secara benar. Di sisi lain diperoleh nilai *False Positive* (FP) sebesar 18 dan *False Negative* (FN) sebesar 243. Jumlah *False Negative* yang relatif tinggi menunjukkan bahwa model masih mengalami keterbatasan dalam mendeteksi *tweet* yang benar-benar termasuk ke dalam kategori bencana. Temuan ini menjadi aspek penting dalam penerapan sistem informasi kebencanaan karena kegagalan mengidentifikasi informasi bencana dapat mengurangi efektivitas penyampaian informasi, memperlambat respons, serta memengaruhi proses pengambilan keputusan. Ringkasan hasil evaluasi performa model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* disajikan pada Tabel 5.

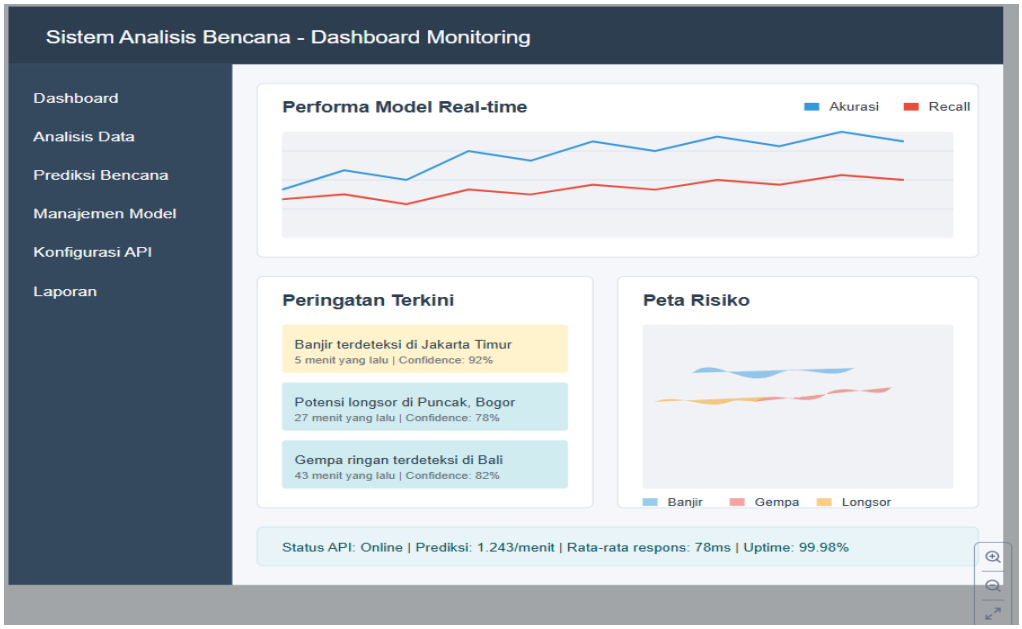
Tabel 5. Hasil Evaluasi Model

Metrik Evaluasi	**Nilai**
Accuracy	89%
Precision	89%
Recall	39%
F1-Score	54%

Hasil evaluasi menunjukkan bahwa model mencapai nilai *accuracy* sebesar 89% yang mengindikasikan kemampuan klasifikasi yang baik secara keseluruhan. Selain itu nilai *precision* sebesar 89% menunjukkan bahwa mayoritas prediksi positif yang dihasilkan model merupakan *tweet* yang benar-benar termasuk dalam kategori bencana sehingga tingkat *False Positive* relatif rendah. Sebaliknya nilai *recall* sebesar 39% mengindikasikan bahwa model belum mampu mengidentifikasi seluruh *tweet* bencana yang terdapat pada *dataset*. Rendahnya nilai *recall* mencerminkan tingginya jumlah *False Negative* yang menunjukkan masih adanya *tweet* bencana yang tidak berhasil dikenali oleh model. Kondisi ini kemungkinan dipengaruhi oleh ketidakseimbangan distribusi kelas (*class imbalance*) serta karakteristik bahasa pada media sosial yang bersifat informal, beragam, dan dinamis. Dibandingkan dengan penelitian sebelumnya yang menggunakan metode *K-Nearest Neighbor* (K-NN) dan *Decision Tree*, model LSTM yang diusulkan menunjukkan peningkatan pada nilai *accuracy* sehingga memiliki potensi yang lebih baik dalam tugas klasifikasi *tweet* bencana.

3.5 Deployment

Tahap terakhir dalam metodologi CRISP-DM adalah *Deployment*. Pada tahap ini model yang telah dibangun diintegrasikan ke dalam *dashboard monitoring* system informasi bencana untuk memudahkan proses visualisasi dan pemantauan *data tweet* secara *real-time*. *Dashboard* dirancang agar dapat digunakan oleh pihak terkait seperti Badan Nasional Penanggulangan Bencana (BNPB), Badan Meteorologi Klimatologi dan Geofisika (BMKG), dan bahkan masyarakat dalam memperoleh informasi terkait bencana alam secara lebih cepat dan terstruktur. Hasilnya dari implementasi sistem pada tahap ini terdapat dalam Gambar 7.



Gambar 7. *Dashboard Monitoring*

Implementasi *dashboard monitoring* diharapkan dapat mendukung proses mitigasi dan tanggap darurat bencana melalui pemanfaatan data media sosial sebagai sumber informasi tambahan. Selain itu integrasi model klasifikasi berbasis LSTM pada sistem *monitoring* memungkinkan proses identifikasi *tweet* bencana dilakukan secara otomatis sehingga dapat membantu meningkatkan efektivitas pengelolaan informasi kebencanaan.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma *Long Short-Term Memory* (LSTM) dengan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk melakukan klasifikasi *tweet* bencana alam gunung meletus pada *platform* Twitter (X). Hasil evaluasi menunjukkan bahwa model mencapai nilai *accuracy* sebesar 89%, *precision* sebesar 89%, *recall* sebesar 39%, dan *F1-score* sebesar 54%. Performa tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam melakukan klasifikasi secara keseluruhan serta menghasilkan tingkat *False Positive* yang rendah. Meskipun demikian nilai *recall* yang relatif rendah mengindikasikan bahwa kemampuan model dalam mengidentifikasi *tweet* bencana masih perlu ditingkatkan. Temuan ini menunjukkan bahwa pengembangan lebih lanjut masih diperlukan untuk meningkatkan sensitivitas model sehingga sistem informasi kebencanaan berbasis media sosial dapat memberikan dukungan yang lebih efektif terhadap proses deteksi dini, respons, dan pengambilan keputusan pada saat terjadi bencana.

Meskipun demikian penelitian ini menunjukkan bahwa pemanfaatan algoritma LSTM dan data dari media sosial memiliki potensi untuk mendukung pengembangan sistem *monitoring* informasi kebencanaan secara otomatis. Penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data, optimasi *hyperparameter*, serta membandingkan performa model LSTM dengan model berbasis *transformer* seperti IndoBERT atau BERT untuk meningkatkan kemampuan deteksi *tweet* bencana secara lebih optimal.

Daftar Pustaka

Airlangga, G. (2024). Comparative Analysis of Machine Learning Models for Real-Time Disaster Tweet Classification : Enhancing Emergency Response with Social Media Analytics. *Brilliance: Research of Artificial Intelligence*, *4*(1), 25–31. https://doi.org/10.47709/brilliance.v4i1.3669

Ashari, Y., & Suhendar, A. (2024). Implementasi Algoritma Long Short-Term Memory (LSTM) untuk Memprediksi Harga Beras di Jawa Tengah Berdasarkan Cuaca. *Djtechno : Jurnal Teknologi Informasi*, *5*(3), 624–636. https://doi.org/10.46576/djtechno

BNPB. (2025). *Data Bencana Indonesia 2024* (A. Muhari, T. Harjito, F. Irawan, & A. C. Utomo (eds.); 3rd ed). Pusat Data Informasi dan Komunikasi Kebencanaan Badan Nasional Penanggulangan Bencana. https://bnpb.go.id/storage/app/media/Buletin Info Bencana/Buku Data Bencana 2024/20250613_Buku Data 2024.pdf

Christanto, F. W., Handayani, S., Handayani, T., & Dewi, C. (2025). APRS and SSTV Technology for Audiovisual Data Transmission in Internet Blank Spot Areas to Increase the Effectiveness of SAR Activities. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, *11*(1), 1–12. https://doi.org/10.26594/register.v11i1.3205

Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi (JATI)*, *12*(1), 64–77. https://doi.org/10.34010/jati.v12i1.6674

Fabillah, D., Auliarahmi, R., Setiarini, S. D., & Gelar, T. (2023). The Investigation of Convolution Layer Structure on BERT-C-LSTM for Topic Classification of Indonesian News Headlines. *Journal of Software Engineering, Information and Communication Technology (SEICT)*, *4*(2), 105–116. https://doi.org/10.17509/seict.v4i2.63742

Fauzianto, R. A., & Supatman. (2023). Analisis Sentimen Opini Masyarakat Terhadap Tech Winter pada Twitter Menggunakan Natural Language Processing. *Jurnal Syntax Admiration*, *4*(9), 1577–1585. https://doi.org/10.46799/jsa.v3i9.909

Hidayat, F. N., S, W. S. J., & Idhom, M. (2026). A Deep Learning Approach Using Bidirectional-LSTM and Word2Vec for Fake News Classification. *Bit-Tech (Binary Digital - Technology)*, *8*(3), 3531–3541. https://doi.org/10.32877/bt.v8i3.3575

Hidayati, N., Suntoro, J., & Setiaji, G. G. (2021). Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM. *Jurnal Sains Dan Informatika*, *7*(2), 117–126. https://doi.org/10.34128/jsi.v7i2.313

Husada, I. N., & Toba, H. (2020). Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2), 400–413. https://doi.org/10.28932/jutisi.v6i2.2743

Kuz, M. V, Lazarovych, I. M., Kozlenko, M. I., Pikuliak, M. V, & Kvasniuk, A. D. (2025). Research on a Hybrid LSTM-CNN-Attention Model for Text-Based Web Content

Classification. *Radio Electronics, Computer Science, Control*, *4*, 105–115. https://doi.org/10.15588/1607-3274-2025-4-10

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N. J. A. H., Ramirez-Quintana, M. J., & Flach, P. A. (2021). CRISP-DM Twenty Years Later : From Data Mining Processes to Data Science Trajectories. *International Conference on Big Data and Education 2019*, 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Minaee, S., Kalchbrenner, N. A. L., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning – based Text Classification : A Comprehensive. *ACM Computing Surveys*, *54*(3), 62. https://doi.org/10.1145/3439726

Mubarak, R., Hanafi, M., & Sasongko, D. (2024). Komparasi Performa Naive Bayes Gaussian dan K-NN Untuk Prediksi Kelulusan Mahasiswa dengan CRISP-DM. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(6), 2982–2991. https://doi.org/10.30865/klik.v4i6.1924

Nofiyanti, E., & Haryanto, E. M. O. N. (2021). Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah Sinus (JIS)*, *19*(2), 17–26. https://doi.org/10.30646/sinus.v19i2.563

Opitz, J. (2024). A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice. *Transactions of the Association for Computational Linguistics*, *12*, 820–836. https://doi.org/10.1162/tacl a 00675

Pandunata, P., Winarno, K. T., Nugroho, Nurdiansyah, Y., & Maidah, N. El. (2022). Analisis Sentimen Opini Publik Terhadap Program Vaksinasi Covid-19 Di Indonesia pada Twitter Menggunakan Metode Naive Bayes Classifier. *INFORMAL: Informatics Journal*, *7*(3), 246–257. https://doi.org/10.19184/isj.v7i3.34930

Qutaishat, D., & Li, S. (2026). A Systematic Review and Comparative Analysis of Deep Learning Models for Twitter/X-based Traffic Event Detection. *International Journal of Digital Earth*, *19*(1), 2604977. https://doi.org/10.1080/17538947.2025.2604977

Rainio, O., Teuho, J., & Klen, R. (2024). Evaluation Metrics and Statistical Tests for Machine Learning. *Scientific Reports*, *14*(6086), 1–14. https://doi.org/10.1038/s41598-024-56706-x

Shahputri, V. A., & Yamasari, Y. (2023). Analisis Sentimen Mengenai Pasca Bencana Alam Menggunakan Metode K-Nearest Neighbor (K-NN) dan Decision Tree. *Journal of Informatics and Computer Science (JINACS)*, *05*(03), 377–388. https://doi.org/10.26740/jinacs.v5n03.p377-388

Stepanenko, V. (2020). *Disaster Tweets*. Kaggle. https://www.kaggle.com/datasets/vstepanenko/disaster-tweets

Sulthan, M. B., Wahyudi, I., & Suhartini, L. (2021). Analisis Sentimen pada Tweet Bencana Alam Menggunakan Deep Neural Network dan Information Gain. *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)*, *2*(2), 65–71. https://doi.org/10.31102/jatim.v2i2.1273

Wahyuni, T., & Susanti, D. (2023). Analisis Potensi Bencana Alam Tanah Longsor Kabupaten Majalengka Menggunakan Algoritma Naïve Bayes Classifier. *INFOTECH Journal*, *9*(2), 299–306. https://doi.org/10.31949/infotech.v9i2.5645