



**KLASIFIKASI PESAN BENCANA ALAM BERBAHASA INDONESIA  
DI MEDIA SOSIAL MENGGUNAKAN INDOBERT DAN  
MULTILINGUAL BERT**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**YASMIN DWI SAFITRI**

**NIM 2111016220005**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**SEPTEMBER 2025**



**KLASIFIKASI PESAN BENCANA ALAM BERBAHASA INDONESIA DI  
MEDIA SOSIAL MENGGUNAKAN INDOBERT DAN MULTILINGUAL BERT**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**YASMIN DWI SAFITRI**

**NIM 2111016220005**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**SEPTEMBER 2025**

## SKRIPSI

### KLASIFIKASI PESAN BENCANA ALAM BERBAHASA INDONESIA DI MEDIA SOSIAL MENGUNAKAN INDOBERT DAN MULTILINGUAL BERT

Oleh:

**Yasmin Dwi Safitri**

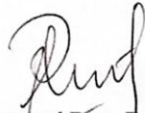
**NIM. 2111016220005**

Telah dipertahankan di depan Dosen Penguji pada tanggal 15 September 2025.

Susunan Dosen Penguji:

**Pembimbing I**

**Dosen Penguji I**



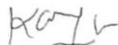
Mohammad Reza Faisal, S.Si., S.T., M.T., Ph.D. Triando Hamonangan Saragih, S.Kom., M.Kom.

NIP. 197612202008121001

NIP. 199308242019031012

**Pembimbing II**

**Dosen Penguji II**



Dwi Kartini, S.Kom., M.Kom.

Friska Abadi, S.Kom., M.Kom.

NIP. 198704212012122003

NIP. 198809132023211010

Banjarbaru, 18 September 2025

**Koordinator Program Studi Ilmu Komputer**



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

## SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 15 September 2025

Yasng Menyatakan,



Yasmin Dwi Safitri

NIM. 2111016220005

## ABSTRAK

### **KLASIFIKASI PESAN BENCANA ALAM BERBAHASA INDONESIA DI MEDIA SOSIAL MENGGUNAKAN INDOBERT DAN MULTILINGUAL BERT**

(Oleh : Yasmin Dwi Safitri; Pembimbing: Mohammad Reza Faisal, S.Si., S.T., M.T., Ph.D. dan Dwi Kartini, S.Kom., M.Kom.; 2025; halaman)

Informasi tentang bencana alam yang disebarkan melalui media sosial dapat menjadi sumber data penting bagi proses mitigasi dan sistem peringatan dini. Platform media sosial, seperti X (sebelumnya dikenal sebagai Twitter), telah menjadi saluran utama untuk menyampaikan informasi secara *real-time*, terutama selama keadaan darurat bencana. Dengan banyaknya teks terkait bencana yang tidak terstruktur yang harus diproses, tantangan utamanya adalah menyaring dan mengklasifikasikan pesan secara akurat ke dalam tiga kategori: saksi mata, bukan saksi mata, dan tidak tahu. Penelitian ini bertujuan untuk membandingkan kinerja empat model pemrosesan bahasa alami berbasis BERT, yaitu IndoBERT, IndoBERT dengan *Masked Language Modeling* (MLM), BERT *Multilingual*, dan BERT *Multilingual* dengan MLM, dalam mengklasifikasikan pesan bencana berbahasa Indonesia. Dataset yang digunakan dalam penelitian ini diperoleh dari penelitian sebelumnya dan data yang tersedia untuk umum di GitHub, yang terdiri dari pesan-pesan berannotasi terkait banjir, gempa bumi, dan kebakaran hutan. Metode yang diterapkan adalah pendekatan pembelajaran mendalam menggunakan teknik *hold-out* dengan rasio 80:20 untuk data pelatihan dan pengujian, dan rasio yang sama diterapkan untuk membagi data pelatihan menjadi subset pelatihan dan validasi, dengan stratifikasi untuk mempertahankan proporsi kelas yang seimbang. Selain itu, variasi ukuran *batch* dieksplorasi untuk mengevaluasi pengaruhnya terhadap stabilitas kinerja model. Hasil penelitian menunjukkan bahwa model IndoBERT mencapai kinerja tertinggi pada dataset banjir dan gempa bumi, dengan akurasi masing-masing 80,67% dan 81,50%. Sementara itu, IndoBERT dengan pra-pelatihan MLM mencatat akurasi tertinggi pada dataset kebakaran hutan, yaitu 88,33%. Secara keseluruhan, IndoBERT menunjukkan kinerja yang paling konsisten dan unggul di seluruh dataset dibandingkan dengan model lainnya. Temuan ini menunjukkan bahwa IndoBERT memiliki kemampuan yang kuat dalam memahami teks terkait bencana Indonesia, dan hasilnya dapat digunakan sebagai dasar untuk mengembangkan sistem klasifikasi otomatis untuk mendukung aplikasi pemantauan bencana dan peringatan dini secara *real-time*.

**Kata kunci:** Pembelajaran Mendalam, Media Sosial, Bencana Alam, IndoBERT, BERT Multibahasa

## ABSTRACT

### **AUTOMATIC ANALYSIS OF NATURAL DISASTER MESSAGES ON SOCIAL MEDIA USING INDOBERT AND MULTILINGUAL BERT**

(Oleh : Yasmin Dwi Safitri; Pembimbing: Mohammad Reza Faisal, S.Si., S.T., M.T., Ph.D. dan Dwi Kartini, S.Kom., M.Kom.; 2025; halaman)

Information about natural disasters disseminated through social media can serve as an important data source for mitigation processes and early warning systems. Social media platforms, such as X (formerly known as Twitter), have become primary channels for conveying real-time information, especially during disaster emergencies. With the large amount of unstructured disaster-related text that must be processed, the main challenge is accurately filtering and classifying messages into three categories: eyewitness, non-eyewitness, and don't know. This research aims to compare the performance of four BERT-based natural language processing models, namely IndoBERT, IndoBERT with Masked Language Modeling (MLM), Multilingual BERT, and Multilingual BERT with MLM, in classifying Indonesian-language disaster messages. The dataset used in this study was obtained from previous research and publicly available data on GitHub, consisting of annotated messages related to floods, earthquakes, and forest fires. The method applied is a deep learning approach using the hold-out technique with an 80:20 ratio for training and testing data, and the same ratio applied to split the training data into training and validation subsets, with stratification to maintain balanced class proportions. In addition, variations in batch size were explored to evaluate their effect on model performance stability. The results show that the IndoBERT model achieved the highest performance on the flood and earthquake datasets, with accuracies of 80.67% and 81.50%, respectively. Meanwhile, IndoBERT with MLM pre-training recorded the highest accuracy on the forest fire dataset, 88.33%. Overall, IndoBERT demonstrated the most consistent and superior performance across datasets compared to the other models. These findings indicate that IndoBERT has strong capabilities in understanding Indonesian disaster-related text, and the results can be used as a foundation for developing automatic classification systems to support real-time disaster monitoring and early warning applications.

**Kata kunci:** Deep Learning, Social Media, Natural Disaster, IndoBERT, Multilingual BERT

## KATA PENGANTAR

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*Automatic Analysis of Natural Disaster Messages on Social Media Using IndoBERT and Multilingual BERT*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada semua pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Allah SWT yang telah memberikan kesehatan, kekuatan, dan kemudahan dalam menyelesaikan jurnal ini.
2. Kepada diri sendiri atas tekad dan semangat yang terus terjaga meskipun menghadapi berbagai tantangan dan hambatan dalam penyusunan jurnal ini.
3. Keluarga tercinta Bapak, Mama, dan Yandi, yang selalu memberikan doa, dukungan, dan kasih sayang.
4. Keluarga besar yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian jurnal ini.
5. Bapak Mohammad Reza Faisal, S.Si, S.T, M.T, Ph.D. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
6. Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen pembimbing pendamping dan Ketua Program Studi Ilmu Komputer FMIPA ULM, yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
7. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
8. Winda Agustina, Rini Hafizah, Elvina Nur Hana, dan Noryasminda selaku sahabat peneliti yang telah memberikan dukungan, kerja sama, serta kontribusi berharga dalam penyusunan penelitian ini.
9. Teman-teman Ryzen 21 serta sahabat-sahabat keluarga Ilmu Komputer yang

telah memberikan dukungan, bantuan, dan doa selama proses penyusunan jurnal ini.

10. Seluruh pihak yang tidak dapat disebutkan satu per satu yang telah berkontribusi dalam membantu penyelesaian jurnal ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Semoga tulisan ini bermanfaat bagi pengembangan ilmu pengetahuan dan memberikan manfaat bagi para pembaca, serta mendapat keridhaan Allah SWT.

Banjarbaru, 15 September 2025



Yasmin Dwi Safitri

NIM. 2111016220005