



**MONITORING OTOMATIS BENCANA ALAM MELALUI PESAN SAKSI
MATA DI MEDIA SOSIAL MENGGUNAKAN VARIAN ALGORITMA
*BERT***

Skripsi

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

Oleh

MUHAMMAD BASHIR HANAFI

NIM 2111016310002

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

JUNI 2025



**MONITORING OTOMATIS BENCANA ALAM MELALUI PESAN SAKSI
MATA DI MEDIA SOSIAL MENGGUNAKAN VARIAN ALGORITMA
*BERT***

Skripsi

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

Oleh

MUHAMMAD BASHIR HANAFI

NIM 2111016310002

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

JUNI 2025

HALAMAN PENGESAHAN

SKRIPSI

MONITORING OTOMATIS BENCANA ALAM MELALUI PESAN SAKSI MATA DI MEDIA SOSIAL MENGGUNAKAN VARIAN ALGORITMA *BERT*

Oleh:

MUHAMMAD BASHIR HANAFI
NIM. 2111016310002

Telah dipertahankan di depan Dosen Penguji pada tanggal 10 Juni 2025.

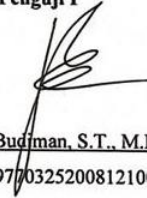
Susunan Dosen Penguji:

Pembimbing I



Mohammad Reza Faisal, S.T., M.T., Ph.D.
NIP. 197612202008121001

Dosen Penguji I



Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Pembimbing II



Friska Abadi, S.Kom., M.Kom.
NIP. 198809132023211010

Dosen Penguji II



Setyo Wahyu Saputro, S.Kom., M.Kom.
NIP. 198808072023211027

Banjarbaru, 24 Juni 2025

Koordinator Program Studi Ilmu Komputer



Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi 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, 2 Juni 2025

Yang Menyatakan,



Muhammad Bashir Hanafi

NIM. 2111016310002

ABSTRAK

MONITORING OTOMATIS BENCANA ALAM MELALUI PESAN SAKSI MATA DI MEDIA SOSIAL MENGGUNAKAN VARIAN ALGORITMA *BERT*

(Oleh: Muhammad Bashir Hanafi; Pembimbing: Mohammad Reza Faisal, S.T., M.T., Ph.D. dan Friska Abadi, S.Kom., M.Kom.; 2025; 77 halaman)

Peningkatan frekuensi dan intensitas bencana alam global mendorong kebutuhan akan sistem pemantauan yang lebih efektif. Media sosial, sebagai sumber informasi *real-time*, menghadapi tantangan dalam memproses volume data yang besar, tidak terstruktur, dan penuh *noise*. Penelitian terdahulu telah memanfaatkan model *machine learning* dan *deep learning* untuk mengatasi monitoring otomatis pesan saksi mata di media sosial, namun masih memiliki keterbatasan. Model *machine learning* masih bergantung pada ekstraksi fitur berbasis frekuensi dan ekstraksi fitur buatan manusia, sehingga masih mengalami kesulitan dalam menangkap kesamaan semantik serta pola sekuensial dalam teks. Model *deep learning* meskipun telah meningkatkan performa secara signifikan, namun masih terdapat keterbatasan dalam memahami semantik dan kurang andal dalam menangani volume data yang besar akibat dari pemrosesan secara sekuensial yang masih sering terjadi di media sosial. Penelitian ini menggunakan model berbasis Transformer dengan varian BERT (BERT, RoBERTa, DistilBERT, ELECTRA, ALBERT) yang diterapkan strategi *pre-training* Masked Language Modeling (MLM) untuk masing-masing model dan optimasi *batch size* dengan tujuan untuk meningkatkan performa. Hasil penelitian menunjukkan bahwa *batch size* 16 memberikan performa tertinggi di seluruh varian BERT dan dataset bencana alam. Model dengan macro F1 score tertinggi terdapat pada model standar BERT (16) sebesar 0,762. Analisis performa berdasarkan jenis bencana menunjukkan macro F1 score sebesar 0,744 (angin), 0,793 (banjir), 0,756 (gempa bumi) dan 0,750 (kebakaran). BERT (16) memiliki performa unggul dibandingkan varian BERT lainnya serta 12 model dari penelitian terdahulu. Penelitian ini memberikan wawasan penting dalam pengembangan model klasifikasi berbasis varian BERT serta mendukung peningkatan efektivitas respons bencana melalui klasifikasi pesan saksi mata bencana alam di media sosial.

Kata Kunci: Klasifikasi, Transformer, Twitter/X, Masked Language Modeling, Batch Size

ABSTRACT

AUTOMATIC MONITORING OF NATURAL DISASTERS THROUGH EYEWITNESS MESSAGES ON SOCIAL MEDIA USING BERT VARIANTS ALGORITHM

(By: Muhammad Bashir Hanafi; Supervisor: Mohammad Reza Faisal, S.T., M.T., Ph.D., and Friska Abadi, S.Kom., M.Kom.; 2025; 77 pages)

The increasing frequency and intensity of global natural disasters has created a demand for more effective monitoring systems. Social media, a real-time information source, presents challenges because of the large volumes of unstructured, noisy data. Earlier studies have applied machine-learning and deep-learning models to automate the monitoring of eyewitness messages on social media, but these models still have shortcomings. Traditional machine-learning approaches depend on frequency-based and handcrafted feature extraction, which hampers their ability to capture semantic similarity and sequential patterns in text. Deep-learning models, while yielding notable performance gains, still struggle with deep semantic understanding and become less reliable when processing large data volumes due to their inherently sequential computation. This issue is pronounced in social-media streams. This study employs transformer-based models using several BERT variants (BERT, RoBERTa, DistilBERT, ELECTRA, and ALBERT). Each model is pre-trained with the Masked Language Modeling (MLM) objective, and batch-size optimization is applied to boost performance. Results show that a batch size of 16 delivers the best overall performance across all BERT variants and disaster datasets. The standard BERT model with batch size 16 achieves the highest macro-F1 score of 0.762. By disaster type, macro-F1 scores reach 0.744 for wind, 0.793 for flood, 0.756 for earthquake, and 0.750 for wildfire. BERT (16) outperforms the other BERT variants and twelve baseline models from prior research. These findings provide important insights for developing BERT-variant-based classification models and support more effective disaster response through improved classification of eyewitness messages on social media.

Keywords: *Classification, Transformer, Twitter/X, Masked Language Modeling, Batch Size*

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Monitoring Otomatis Bencana Alam melalui Pesan Saksi Mata di Media Sosial menggunakan Varian Algoritma *BERT*” untuk memenuhi syarat dalam menyelesaikan pendidikan S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta sahabat, keluarga, dan pengikut beliau hingga *yaumul qiama*.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga tersayang yaitu Bapak Irianto, Almarhumah Ibu Yatmi, Ibu Mulyati, Kak Siti Nurhasanah, Kak Desi Norjannah, dan Mas Dimas yang senantiasa memberikan doa, semangat, dukungan, hingga kepercayaan, sehingga penulis termotivasi untuk mengejar pendidikan setinggi mungkin serta dapat menyelesaikan skripsi ini.
2. Bapak M. Reza Faisal, S.T., M.T., Ph.D. selaku dosen pembimbing utama dan Bapak Friska Abadi, S.Kom., M.Kom. selaku dosen pembimbing pendamping, yang senantiasa membimbing, membantu, dan meluangkan waktu dalam proses penyelesaian skripsi ini.
3. Bapak Irwan Budiman, S.T., M.Kom. selaku Dosen Pembimbing Akademik dan Ibu Dwi Kartini, S.Kom., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
4. Seluruh dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini sangat bermanfaat.

5. Para teman seperjuangan saya, Adis, Yasir, Pita, Nida, Tasya, Nazar, Anissa, Mala, Dzira, dan Tiara, terima kasih telah hadir sebagai pelabuhan semangat dan pengingat bahwa saya tidak pernah berjalan sendirian.
6. Brodie HAFECS, Geng Lolok, Itinergo's People, Tim APEHIPO, dan *mutual* Twitter, terima kasih telah menjadi ruang aman dan tempat saya kembali saat langkah terasa berat. Dukungan kalian menjadi bahan bakar semangat dalam menyelesaikan skripsi ini sebaik mungkin.
7. Keluarga Ryzen'21, terima kasih untuk canda, tawa, perjuangan yang sudah kita lewati bersama, untuk semua kenangan manis yang terukir selama ini. Senang bisa menjadi salah satu bagian dari kehidupan kalian.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh lebih dari sempurna dan penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 2 Juni 2025



Muhammad Bashir Hanafi

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Kajian Terdahulu.....	7
2.2 Landasan Teori.....	11
2.2.1 Media Sosial.....	11
2.2.2 Twitter/X.....	11
2.2.3 Pesan Saksi Mata	11

2.2.4 Natural Language Processing.....	12
2.2.5 Klasifikasi Teks.....	13
2.2.6 Transformer.....	13
2.2.7 Lapisan Encoder.....	15
2.2.8 Lapisan Decoder.....	19
2.2.9 BERT	21
2.2.10 Masked Language Modeling.....	23
2.2.11 RoBERTa.....	23
2.2.12 DistilBERT.....	24
2.2.13 ELECTRA.....	24
2.2.14 ALBERT	25
2.2.15 Batch Size.....	25
2.2.16 Confusion Matrix	25
BAB III METODE PENELITIAN.....	29
3.1 Alat Penelitian.....	29
3.2 Bahan Penelitian.....	29
3.3 Variabel Penelitian.....	29
3.4 Prosedur Penelitian.....	29
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Hasil	34
4.1.1 Pengumpulan Dataset.....	34
4.1.2 Preprocessing	35
4.1.3 Penambahan Token Spesial.....	40
4.1.4 Pelatihan Model MLM.....	41
4.1.5 Klasifikasi	43

4.1.6 Evaluasi	54
4.2 Pembahasan.....	64
BAB V PENUTUP.....	73
5.1 Kesimpulan	73
5.2 Saran.....	73
DAFTAR PUSTAKA	74
LAMPIRAN	78
RIWAYAT PENULIS	92

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	8
Tabel 2. Perancangan Penelitian	10
Tabel 3. Rincian Dataset	31
Tabel 4. Contoh Data Pesan Saksi Mata Bencana Alam Banjir	34
Tabel 5. Contoh Penerapan Eliminasi Spasi Ganda.....	35
Tabel 6. Contoh Penerapan Eliminasi Tanda Baca.....	36
Tabel 7. Contoh Penerapan Eliminasi Angka dan URL	37
Tabel 8. Contoh Penerapan Eliminasi Karakter Non-alfanumerik	38
Tabel 9. Contoh Penerapan Case Folding	39
Tabel 10. Contoh Penerapan Penambahan Token Spesial.....	40
Tabel 11. Contoh Penerapan <i>Masking</i>	41
Tabel 12. Hasil Pelatihan Model MLM	42
Tabel 13. Hyperparameter Model Klasifikasi.....	43
Tabel 14. Hasil Klasifikasi Model BERT	45
Tabel 15. Hasil Klasifikasi Model RoBERTa.....	46
Tabel 16. Hasil Klasifikasi Model DistilBERT	47
Tabel 17. Hasil Klasifikasi Model ELECTRA	48
Tabel 18. Hasil Klasifikasi Model ALBERT.....	49
Tabel 19. Hasil Klasifikasi Model BERT-MLM	50
Tabel 20. Hasil Klasifikasi Model RoBERTa-MLM.....	51
Tabel 21. Hasil Klasifikasi Model DistilBERT-MLM	52
Tabel 22. Hasil Klasifikasi Model ELECTRA-MLM.....	53
Tabel 23. Hasil Klasifikasi Model ALBERT-MLM.....	54
Tabel 24. Evaluasi Direct_eyewitnesses BERT (16) Dataset Angin	55
Tabel 25. Evaluasi Noneyewitnesses BERT (16) Dataset Angin	56
Tabel 26. Evaluasi Don't-know BERT (16) Dataset Angin	56
Tabel 27. Hasil Evaluasi Macro F1-Score BERT (16) Bencana Angin....	57
Tabel 28. Evaluasi Direct_eyewitnesses BERT (16) Dataset Banjir	58

Tabel 29. Evaluasi Noneyewitnesses BERT (16) Dataset Banjir	58
Tabel 30. Hasil Evaluasi Kelas Don't-know BERT (16) Bencana Banjir	59
Tabel 31. Hasil Evaluasi Macro F1-Score BERT (16) Bencana Banjir....	59
Tabel 32. Evaluasi Direct_eyewitnesses BERT (16) Dataset Gempa.....	60
Tabel 33. Evaluasi Noneyewitnesses BERT (16) Dataset Gempa.....	61
Tabel 34. Evaluasi Don't-know BERT (16) Dataset Gempa.....	61
Tabel 35. Evaluasi Macro F1-Score BERT (16) Bencana Gempa.....	61
Tabel 36. Evaluasi Direct_eyewitnesses BERT (16) Dataset Kebakaran.	62
Tabel 37. Evaluasi Noneyewitnesses BERT (16) Dataset Kebakaran.....	63
Tabel 38. Evaluasi Don't-know BERT (16) Dataset Kebakaran	63
Tabel 39. Evaluasi Macro F1-Score BERT (16) Bencana Kebakaran.....	64
Tabel 40. Evaluasi Performa Model BERT	64
Tabel 41. Perbandingan Performa Model	70

DAFTAR GAMBAR

Gambar 1. Arsitektur Transformer.....	14
Gambar 2. Lapisan Encoder.....	15
Gambar 3. Multi-Head Attention.....	16
Gambar 4. Lapisan Decoder.....	19
Gambar 5. Lapisan Decoder.....	20
Gambar 6. Arsitektur Transformer.....	22
Gambar 7. Masked Language Models.....	23
Gambar 8. Multi-class Confusion Matrix.....	26
Gambar 9. Alur Penelitian.....	30
Gambar 10. Grafik Pelatihan Model BERT.....	44
Gambar 11. Grafik Pelatihan Model RoBERTa.....	45
Gambar 12. Grafik Pelatihan Model DistilBERT.....	46
Gambar 13. Grafik Pelatihan Model ELECTRA.....	47
Gambar 14. Grafik Pelatihan Model ALBERT.....	48
Gambar 15. Grafik Pelatihan Model BERT-MLM.....	49
Gambar 16. Grafik Pelatihan Model RoBERTa-MLM.....	50
Gambar 17. Grafik Pelatihan Model DistilBERT-MLM.....	51
Gambar 18. Grafik Pelatihan Model ELECTRA-MLM.....	52
Gambar 19. Grafik Pelatihan Model ALBERT-MLM.....	53
Gambar 20. Confusion Matrix BERT (16) Dataset Angin.....	55
Gambar 21. Confusion Matrix BERT (16) Dataset Bencana Banjir.....	57
Gambar 22. Confusion Matrix BERT (16) Dataset Gempa.....	60
Gambar 23. Confusion Matrix BERT (16) Dataset Kebakaran.....	62
Gambar 24. Rerata Batch Size - Seluruh Model.....	65
Gambar 25. Rerata Batch Size - Seluruh Dataset Bencana Alam.....	66
Gambar 26. Perbandingan Hasil Batch Size – Bencana Alam.....	66
Gambar 27. Perbandingan Hasil Batch Size - Model.....	66
Gambar 28. Performa Seluruh Varian BERT Batch Size 16.....	67

Gambar 29. Performa Seluruh Varian BERT – Dataset Bencana Alam...	69
Gambar 30. Performa BERT (16) – Dataset Bencana Alam	70

DAFTAR LAMPIRAN

Lampiran 1. Pelatihan MLM - Source code import library	79
Lampiran 2. Pelatihan MLM - Source code load dataset.....	79
Lampiran 3. Pelatihan MLM - Source code tokenisasi dan encoding	79
Lampiran 4. Pelatihan MLM - Source code load model dan tokenizer	80
Lampiran 5. Pelatihan MLM - Source code training model	81
Lampiran 6. Pelatihan MLM – Source code save model.....	81
Lampiran 7. Klasifikasi - Source code import library.....	82
Lampiran 8. Klasifikasi - Source code load dataset.....	82
Lampiran 9. Klasifikasi - Source code tokenisasi dan encoding	83
Lampiran 10. Klasifikasi - Source code load model	83
Lampiran 11. Klasifikasi - Source code class Trainer	85
Lampiran 12. Klasifikasi - Source code load dataset dan split data.....	87
Lampiran 13. Klasifikasi - Source code parameter	88
Lampiran 14. Klasifikasi - Source code training	88
Lampiran 15. Klasifikasi - Source code evaluasi.....	88
Lampiran 16. Hasil Model Klasifikasi (Macro F1 Score).....	90
Lampiran 17. Penelitian Terdahulu.....	91