



**KLASIFIKASI PESAN BENCANA ALAM DARI SOSIAL MEDIA
DENGAN ALGORITMA CNN-LSTM**

Skripsi

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

**Oleh
Muhammad Hamdani
NIM. 1811016110017**

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

JUNI 2025



**KLASIFIKASI PESAN BENCANA ALAM DARI SOSIAL MEDIA
DENGAN ALGORITMA CNN-LSTM**

Skripsi

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

**Oleh
Muhammad Hamdani
NIM. 1811016110017**

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

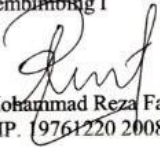
JUNI 2025

SKRIPSI
KLASIFIKASI PESAN BENCANA ALAM DARI SOSIAL MEDIA
DENGAN ALGORITMA CNN-LSTM


Oleh
Muhammad Hamdani
NIM. 1811016110017

Telah dipertahankan di depan Dosen Penguji pada tanggal 26 Juni 2025
Susunan Dosen Penguji :

Pembimbing I


Muhammad Reza Faisal, S.T., M.T., PhD.
NIP. 19761220 200812 1 001

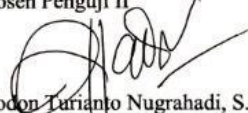
Dosen Penguji I


Radityo Adi Nugroho, S.T, M.Kom
NIP. 19821204200801 1 006

Pembimbing II


Rudy Herjono, S.Kom, M.Kom
NIP. 19880925202203 1 003

Dosen Penguji II


Dodon Turiyanto Nugrahadi, S.Kom, M.Eng
NIP. 19800112200912 1 002

Banjarbaru, Juli 2025

Koordinator Program Studi Ilmu Komputer,


Dwi Karlini, S.Kom., M.Kom.
NIP. 19870421201212 2 002

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, 04 Juli 2025



Muhammad Hamdani
NIM. 1811016110017

ABSTRAK

KLASIFIKASI PESAN BENCANA ALAM DARI SOSIAL MEDIA DENGAN ALGORITMA CNN-LSTM (Oleh : Muhammad Hamdani; Pembimbing: Mohammad Reza Faisal, S.T., M.T., Ph.D dan Rudy Herteno, S.Kom, M.Kom.; 2025; 95 halaman)

Informasi bencana alam yang cepat dan akurat sangat penting untuk mendukung proses penanganan darurat dan penyebaran informasi kepada masyarakat. Media sosial seperti Twitter menjadi sumber informasi potensial karena kecepatan dan keterbukaannya dalam menyampaikan kejadian secara langsung dari masyarakat. Salah satu tantangan utama dalam pemanfaatannya adalah klasifikasi jenis informasi berdasarkan sumber, khususnya untuk membedakan pesan dari saksi mata, non-saksi mata, dan tidak diketahui pada data Twitter terkait tiga jenis bencana alam: banjir, gempa bumi, dan kebakaran hutan. Penelitian ini mengembangkan model klasifikasi menggunakan dua teknik *word embedding*, yaitu Word2Vec dan FastText, yang dikombinasikan dengan arsitektur deep learning CNN - LSTM dalam konfigurasi single-layer dan multi-layer. Hasil evaluasi menunjukkan bahwa model dengan embedding FastText dan arsitektur multi-layer memberikan performa terbaik dalam mengidentifikasi pesan dari saksi mata, khususnya pada data gempa dan kebakaran hutan, dengan sensitivitas mencapai 79,50% dan spesifisitas 87,50% pada kelas *eyewitness*. Sebaliknya, model Word2Vec single-layer menunjukkan performa yang lebih rendah, misalnya pada data banjir dengan spesifisitas hanya 72,50% dan sensitivitas 88,00%, serta pada gempa dengan sensitivitas 82,50% dan spesifisitas 82,25%. Temuan ini menunjukkan bahwa pemilihan embedding dan arsitektur model berpengaruh signifikan terhadap kemampuan klasifikasi, di mana kombinasi FastText dengan arsitektur multi-layer unggul dalam menjaga keseimbangan sensitivitas dan spesifisitas pada pesan dari saksi mata.

ABSTRACT

CLASSIFICATION OF NATURAL DISASTER MESSAGES FROM SOCIAL MEDIA USING CNN-LSTM ALGORITHM (By : Muhammad Hamdani; Supervisor: Mohammad Reza Faisal, S.T., M.T., Ph.D and Rudy Herteno, S.Kom, M.Kom.; 2025; 95 pages)

Rapid and accurate disaster information is crucial to support emergency response efforts and the dissemination of information to the public. Social media platforms such as Twitter have become a potential source of real-time information due to their speed and openness in conveying events directly from the public. One of the main challenges in utilizing this data is classifying the type of information based on its source, particularly in distinguishing messages from *eyewitnesses*, *non-eyewitnesses*, and unknown sources in Twitter data related to three types of natural disasters: floods, earthquakes, and wildfires. This study develops a classification model using two *word embedding* techniques, namely Word2Vec and FastText, combined with a CNN - LSTM deep learning architecture in both single-layer and multi-layer configurations. The evaluation results indicate that models utilizing FastText embeddings and a multi-layer architecture achieve the best performance in identifying *eyewitness* messages, particularly for earthquake and wildfire data, with sensitivity reaching 79.50% and specificity at 87.50% for the *eyewitness* class. In contrast, the Word2Vec single-layer model shows lower performance, for instance, in flood data with specificity of only 72.50% and sensitivity of 88.00%, and in earthquake data with sensitivity of 82.50% and specificity of 82.25%. These findings demonstrate that the choice of embedding and model architecture significantly affects classification performance, with the combination of FastText and a multi-layer architecture outperforming in maintaining a balance between sensitivity and specificity in classifying *eyewitness* messages.

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Pesan Bencana Alam dari Sosial Media dengan Algoritma CNN-LSTM” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad ﷺ beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiamah*.

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


1. Diri saya sendiri yang tidak pernah patah semangat walaupun banyak menemui kesulitan baik disebabkan oleh diri sendiri maupun dari hal lain.
2. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
3. Bapak Mohammad Reza Faisal, S.T., M.T., Ph.D selaku dosen pembimbing utama dan pembimbing akademik yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Rudy Herteno, S.Kom, M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
5. Ibu Dwi Kartini S.Kom, M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
6. Bapak Radityo Adi Nugroho, S.T, M.Kom dan Dodon Turianto Nugrahadi, S.Kom, M.Eng selaku dosen penguji yang telah memberikan masukan dalam skripsi ini.
7. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.

8. Teman-teman keluarga Ilmu Komputer angkatan 2018 yang telah lulus lebih dahulu dan juga yang belum sehingga memberikan motivasi untuk menyelesaikan skripsi ini dan juga tempat bertanya disaat kebingungan.
9. 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 dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik 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, 04 Juli 2025



Muhammad Hamdani

DAFTAR ISI

SKRIPSI.....	iii
PERNYATAAN.....	iii
ABSTRAK	v
ABSTRACT	vi
PRAKATA	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu	5
2.2 Twitter	6
2.3 Social Network Sensor	7
2.4 Natural Language Processing (NLP)	7
2.5 Teks Mining	8
2.6 <i>Word Embedding</i>	9
2.7 Word2Vec	9
2.8 Fast Text	11
2.9 Convolutional Neural Network (CNN).....	11

2.9.1	<i>Convolution Layer</i>	12
2.9.2	Pooling Layer.....	13
2.9.3	Fully Connected Layer.....	14
2.10	Long Short Term Memory.....	14
2.11	Convolutuional Neural Network - Long Short Term Memory (CNN-LSTM).....	17
2.12	Evaluasi Klasifikasi.....	17
2.13	Keaslian Penelitian.....	20
BAB III METODE PENELITIAN.....		24
3.1	Alat Penelitian.....	24
3.2	Bahan Penelitian.....	24
3.3	Variabel Penelitian.....	24
3.4	Prosedur Penelitian.....	24
3.4.1	Pengumpulan Data.....	25
3.4.2	<i>Preprocessing</i>	26
3.4.3	Pembuatan Model <i>Word Embedding</i>	27
3.4.4	Pembagian Data.....	27
3.4.5	Pembuatan Model Klasifikasi.....	27
3.4.6	Evaluasi.....	28
BAB IV HASIL DAN PEMBAHASAN.....		29
4.1	Hasil.....	29
4.1.1	Pengumpulan Data.....	29
4.1.2	<i>Preprocessing</i> Data.....	30
4.1.3	Pembuatan Model <i>Word Embedding</i>	31
4.1.4	Pembobotan dengan Word2Vec dan FastText.....	33
4.1.5	Pembuatan Model Klasifikasi.....	34
4.1.6	Klasifikasi dan Evaluasi.....	37

4.2	Pembahasan.....	50
BAB V PENUTUP.....		57
5.1	Kesimpulan.....	57
5.2	Saran	57
DAFTAR PUSTAKA		58

DAFTAR TABEL

Tabel 1. Perbandingan data berdasarkan label	3
Tabel 2. <i>Confusion Matrix</i> Multiclass	18
Tabel 3. Keaslian Penelitian.....	21
Tabel 4. Contoh data bencana banjir	25
Tabel 5. Contoh data bencana gempa	26
Tabel 6. Contoh data bencana kebakaran hutan	26
Tabel 7. Contoh data bencana banjir	29
Tabel 8. Contoh data bencana gempa	29
Tabel 9. Contoh data bencana kebakaran hutan	30
Tabel 10. Cleansing dan Case folding pada data	30
Tabel 11. Model pretrained Word2Vec	32
Tabel 12. Model pretrained Fast Text	33
Tabel 13. Hasil Kinerja CNN-LSTM data Banjir Word2Vec	40
Tabel 14. Hasil Kinerja CNN-LSTM data Gempa Word2Vec	43
Tabel 15. Hasil Kinerja CNN-LSTM data Kebakaran Hutan Word2Vec	46
Tabel 16. Hasil Kinerja CNN-LSTM data Banjir FastText	47
Tabel 17. Hasil Kinerja CNN-LSTM data Gempa FastText	48
Tabel 18. Hasil Kinerja CNN-LSTM data Kebakaran Hutan FastText	49
Tabel 19. Nilai sensitivity model CNN-LSTM berbasis Word2Vec	52
Tabel 20. Nilai sensitivity model CNN-LSTM berbasis FastText.....	53
Tabel 21. Nilai specificity model CNN-LSTM berbasis Word2Vec	54
Tabel 22. Nilai specificity model CNN-LSTM berbasis FastText	55

DAFTAR GAMBAR

Gambar 1. Model Skip-Gram.....	10
Gambar 2. Arsitektur standar dari CNN yang terdiri dari lapisan konvolusi, lapisan pooling, dan lapisan dense (Hu et al., 2015).....	12
Gambar 3. Proses yang terjadi pada lapisan konvolusional.....	13
Gambar 4. Proses <i>pooling</i> pada CNN.....	14
Gambar 5. Jaringan arsitektur LSTM.....	15
Gambar 6. Contoh arsitektur CNN-LSTM.....	17
Gambar 7. Alur Penelitian.....	25
Gambar 8. Pembuatan <i>Word Embedding</i>	32
Gambar 9. Model multi layer CNN – LSTM menggunakan data banjir.....	38
Gambar 10. Model single layer CNN – LSTM menggunakan data banjir.....	39
Gambar 11. <i>Confusion Matrix</i> data banjir Word2Vec.....	40
Gambar 12. Model multi layer CNN – LSTM menggunakan data gempa.....	41
Gambar 13. Model single layer CNN – LSTM menggunakan data gempa.....	42
Gambar 14. <i>Confusion matrix</i> data Gempa Word2Vec.....	43
Gambar 15. Model multi layer CNN – LSTM menggunakan data kebakaran hutan.....	44
Gambar 16. Model single layer CNN – LSTM menggunakan data kebakaran hutan.....	45
Gambar 17. <i>Confusion Matrix</i> data Kebakaran Hutan Word2Vec.....	46
Gambar 18. <i>Confusion Matrix</i> data Banjir FastText.....	47
Gambar 19. <i>Confusion Matrix</i> data Gempa FastText.....	48
Gambar 20. <i>Confusion Matrix</i> data Kebakaran Hutan FastText.....	49
Gambar 21. Akurasi pengujian dari data banjir, gempa, dan kebakaran hutan dengan <i>word embedding</i> word2vec pada model multi layer dan single layer.....	50
Gambar 22. Akurasi pengujian dari data banjir, gempa, dan kebakaran hutan dengan <i>word embedding</i> fast text pada model multi layer dan single layer.....	51
Gambar 23. Grafik perbandingan sensitivity dengan data banjir, gempa dan kebakaran hutan dari Word2Vec pada model multi layer dan single layer.....	52
Gambar 24. Grafik perbandingan sensitivity dengan data banjir, gempa dan kebakaran hutan dari FastText pada model multi layer dan single layer.....	53
Gambar 25. Grafik perbandingan specificity dengan data banjir, gempa dan kebakaran hutan dari Word2Vec pada model multi layer dan single layer.....	54
Gambar 26. Grafik perbandingan specificity dengan data banjir, gempa dan kebakaran hutan dari FastText pada model multi layer dan single layer.....	54

DAFTAR LAMPIRAN