

PENGARUH TAHAPAN *PREPROCESSING* TERHADAP PERFORMA MODEL INDOBERT DAN INDOBERTWEET PADA ANALISIS SENTIMEN BANJIR ACEH 2025

NAIYA DZIL IZZATI



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pengaruh Tahapan *Preprocessing* terhadap Performa Model IndoBERT dan IndoBERTweet pada Analisis Sentimen Banjir Aceh 2025” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Naiya Dzil Izzati
G1401221065



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

NAIYA DZIL IZZATI. Pengaruh Tahapan *Preprocessing* terhadap Performa Model IndoBERT dan IndoBERTweet pada Analisis Sentimen Banjir Aceh 2025. Dibimbing oleh FARIT MOCHAMAD AFENDI dan MOHAMMAD MASJKUR.

Banjir merupakan salah satu bencana hidrometeorologi yang paling sering terjadi di Indonesia dan sering menjadi topik pembahasan masyarakat di media sosial. Penelitian ini bertujuan menganalisis pengaruh tahapan *preprocessing* terhadap performa model IndoBERT dan IndoBERTweet pada klasifikasi sentimen terkait banjir Aceh. Data penelitian terdiri atas 5403 cuitan yang diberi label secara manual ke dalam tiga kelas sentimen, yaitu negatif, netral, dan positif. Penelitian menerapkan tiga skenario *preprocessing* dengan tingkat kompleksitas yang berbeda, kemudian performa model dievaluasi menggunakan *balanced accuracy* dan *F1-score macro*. Hasil penelitian menunjukkan bahwa IndoBERT mencapai performa terbaik pada skenario *preprocessing* minimal dengan nilai *balanced accuracy* sebesar 0,8575 dan *F1-score macro* sebesar 0,8428. Sedangkan, IndoBERTweet memperoleh performa terbaik dengan nilai *balanced accuracy* sebesar 0,8677 *F1-score macro* sebesar 0,8435, sekaligus menjadi model dengan performa tertinggi pada penelitian ini. IndoBERTweet menunjukkan performa yang lebih konsisten dibandingkan IndoBERT pada seluruh skenario *preprocessing*. Sebaliknya, performa IndoBERT cenderung menurun seiring bertambahnya tahapan *preprocessing*. Temuan ini menunjukkan bahwa pengaruh *preprocessing* berbeda pada setiap model sehingga pemilihan tahapan *preprocessing* perlu disesuaikan dengan karakteristik model dan data yang digunakan.

Kata kunci: analisis sentimen, banjir, IndoBERT, IndoBERTweet, *preprocessing*



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRACT

NAIYA DZIL IZZATI. A Comparative Study of IndoBERT and IndoBERTweet Model under Different *Preprocessing* Scenarios for Sentiment Analysis of the 2025 Aceh Flood. Supervised by FARIT MOCHAMAD AFENDI and MOHAMMAD MASJKUR.

Floods are one of the most frequent hydrometeorological disasters in Indonesia and are often discussed on social media. This study aimed to analyze the effect of preprocessing on the performance of IndoBERT and IndoBERTweet models for sentiment classification related to the 2025 Aceh flood. The dataset consisted of 5,403 tweets that were manually labeled into three sentiment classes: negative, neutral, and positive. Three preprocessing scenarios with different levels of complexity were applied, and model performance was evaluated using balanced accuracy and macro F1-score. The results showed that IndoBERT achieved its best performance under the minimal preprocessing scenario, with a balanced accuracy of 0.8575 and a macro F1-score of 0.8428. Meanwhile, IndoBERTweet achieved its best performance under a more complex preprocessing scenario, with a balanced accuracy of 0.8677 and a macro F1-score of 0.8435, making it the best-performing model in this study. In general, IndoBERTweet showed more consistent performance than IndoBERT across all preprocessing scenarios. In contrast, the performance of IndoBERT decreased as the preprocessing stages became more complex. These findings indicate that the effect of preprocessing differs across models, and therefore preprocessing strategies should be adjusted to the characteristics of the model and the data used.

Keywords: Aceh flood, IndoBERT, IndoBERTweet, preprocessing, sentiment.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

IPB University



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026¹
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

PENGARUH TAHAPAN *PREPROCESSING* TERHADAP PERFORMA MODEL INDOBERT DAN INDOBERTWEET PADA ANALISIS SENTIMEN BANJIR ACEH 2025

NAIYA DZIL IZZATI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika
pada Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
1. Dr. Ir. Erfiani, M.Si.



Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Pengaruh Tahapan *Preprocessing* terhadap Performa Model
IndoBERT dan IndoBERTweet pada Sentimen Banjir Aceh 2025

Nama : Naiya Dzil Izzati

NIM : G1401221065

Disetujui oleh

Pembimbing 1:

Dr. Farit Mochamad Afendi, S.Si., M.Si.



Pembimbing 2:

Mohammad Masjkur, M.S.

Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si, M.Si

NIP 19780411 200501 1002

Tanggal Ujian:
17 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Februari 2026 sampai bulan Juni 2026 ini ialah analisis sentimen, dengan judul “Pengaruh Tahapan *Preprocessing* terhadap Performa Model IndoBERT dan IndoBERTweet pada Sentimen Banjir Aceh 2025”.

Terima kasih penulis ucapkan kepada seluruh pihak yang telah memberikan bimbingan, dukungan, serta doa dalam penyusunan karya ilmiah ini, di antaranya:

1. Ibu Anny Setiawati dan Alm. Bapak Sutekno sebagai orang tua penulis yang selalu memberikan doa, semangat, dukungan, dan kasih sayang kepada penulis.
2. Bapak Dr. Farit Mochamad Afendi, S.Si., M.Si., dan Bapak Ir. Mohammad Masjkur, M.S. selaku dosen pembimbing yang selalu memberikan saran serta arahan pada penulis dalam menyelesaikan penulisan karya ilmiah ini dengan baik.
3. Ibu Dr. Ir. Erfiani, M.Si. selaku dosen penguji yang telah memberikan banyak saran dan masukan kepada penulis dalam menyelesaikan penulisan karya ilmiah ini.
4. Ibu Dr. Utami Dyah Syafitri, S.Si., M.Si., dan Ibu Eka Putri Nur Utami, S.Stat, M.Si., selaku moderator kolokium dan seminar hasil yang telah memberikan saran serta masukkan untuk perbaikan karya ilmiah ini.
5. Seluruh dosen dan keluarga besar Prodi Statistika dan Sains Data SSMI IPB yang telah memberikan ilmu dan pembelajaran yang bermanfaat bagi penulis, serta staf Tata Usaha Prodi Statistika dan Sains Data SSMI yang telah membantu penulis dalam proses administrasi perkuliahan.
6. Zalfa, Jasmita, Aura, Aurel, Keyzha, Yumna, Haviz, Janira, Tazkya, Nazira, Fahriza, Fuzi, dan Sarah selaku sahabat penulis yang selalu ada untuk membersamai sejak awal kuliah, memberikan dukungan dan doa.
7. Seluruh teman-teman Statistika 59 dan seluruh pihak yang telah memberikan kontribusi dan kebaikan mereka, meskipun tidak dapat disebutkan satu per satu, sehingga penulis berhasil menyelesaikan karya ilmiah ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2026

Naiya Dzil Izzati

DAFTAR ISI

DAFTAR TABEL	xvi
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN	xvi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	5
2.1 Analisis Sentimen	5
2.2 IndoBERT	5
2.3 <i>Class Weight</i>	6
2.4 Evaluasi Performa Model	7
III METODE	9
3.1 Data	9
3.2 Prosedur Kerja	9
IV HASIL DAN PEMBAHASAN	13
4.1 Pelabelan Data	13
4.2 <i>Preprocessing</i>	14
4.3 Visualisasi Data	15
4.4 Pemodelan IndoBERT dan IndoBERTweet	17
4.5 Evaluasi Model	18
4.6 Perbandingan Performa Model	22
4.7 Analisis Hasil Prediksi	23
V SIMPULAN DAN SARAN	27
5.1 Simpulan	27
5.2 Saran	27
DAFTAR PUSTAKA	29
LAMPIRAN	31
RIWAYAT HIDUP	37



DAFTAR TABEL

1	<i>Confusion matrix multiclass</i>	7
2	Contoh data hasil <i>scrapping</i>	9
3	Kriteria pelabelan data	10
4	Kombinasi <i>hyperparameter</i> yang digunakan	12
5	Contoh data hasil pelabelan	13
6	Contoh data hasil <i>preprocessing</i>	15
7	Hasil perhitungan <i>class weight</i>	17
8	Hasil <i>tuning hyperparameter</i> pada data validasi model IndoBERT	18
9	Hasil <i>tuning hyperparameter</i> pada data validasi model IndoBERTweet	18
10	<i>Classification report</i> dari model IndoBERT	19
11	<i>Classification report</i> dari model IndoBERTweet	21
12	Perbandingan performa model pada data uji	22
13	Contoh hasil prediksi IndoBERT	24
14	Contoh Hasil Prediksi IndoBERTweet	24

DAFTAR GAMBAR

15	Arsitektur model IndoBERT (Anadra <i>et al.</i> 2025)	6
16	<i>Flowchart</i> prosedur analisis	12
17	<i>Pie chart</i> distribusi data setiap kelas sentimen	14
18	Visualisasi <i>word cloud</i> (a) skenario 1, (b) skenario 2, (c) skenario 3	16
19	Visualisasi <i>word cloud</i> skenario 3 (a) netral, (b) negatif, (c) positif	16
20	Visualisasi <i>top word</i> skenario 3 (a) negatif, (b) netral, (c) positif	17
21	<i>Confusion matrix</i> model IndoBERT (a) skenario 1, (b) skenario 2, dan (c) skenario 3	19
22	<i>Confusion matrix</i> model IndoBERTweet (a) skenario 1, (b) skenario 2, dan (c) skenario 3	21

DAFTAR LAMPIRAN

23	Hasil kombinasi pengujian <i>hyperparameter</i> model IndoBERT skenario 1 pada data validasi	33
24	Hasil kombinasi pengujian <i>hyperparameter</i> model IndoBERT skenario 2 pada data validasi	33
25	Hasil kombinasi pengujian <i>hyperparameter</i> model IndoBERT skenario 2 pada data validasi	34
26	Hasil kombinasi pengujian <i>hyperparameter</i> model IndoBERT skenario 2 pada data validasi	34

27	Hasil kombinasi pengujian hyperparameter model IndoBERT skenario 2 pada data validasi	35
28	Hasil kombinasi pengujian hyperparameter model IndoBERT skenario 2 pada data validasi	35

@Hak cipta milik IPB University

IPB University

