

**PENGARUH NORMALISASI TEKS TERHADAP ANALISIS SENTIMEN
BERBASIS INDOBERTWEET PADA TWEET TENTANG RUU TNI**

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh sarjana komputer



ISNAENI ZAKIYAH

1212001025

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE**

2025

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dari semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar

Nama : Isnaeni Zakiyah
NIM : 1212001025
Tanggal : 02 September 2025

Tanda Tangan : 

HALAMAN PENGESAHAN

Nama : Isnaeni Zakiyah
NIM : 1212001025
Program Studi : Teknik Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Tugas Akhir : **PENGARUH NORMALISASI TEKS TERHADAP
ANALISIS SENTIMEN BERBASIS INDOBERTWEET PADA TWEET
TENTANG RUU TNI**

Telah dinyatakan lulus sidang tugas akhir pada Kamis, 28 Agustus 2025,
di hadapan dosen pembimbing dan dosen penguji dibawah ini:

Jakarta, 30 Agustus 2025

Pembimbing 1



Ir. Iwan Adhicandra, S.T., M.Sc., Ph.D., SMIEEE

Pembahas 1



Albert Arapenta Sembiring, S.T., M.T

Pembahas 2



Prof. Dr. Hoga Saragih, S.T., M.T.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pengaruh Normalisasi Teks Terhadap Analisis Sentimen Berbasis IndoBERTweet pada Tweet Tentang Ruu Tni”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi Teknik Informatika.

Penelitian ini bertujuan untuk menganalisis sentimen dari cuitan masyarakat terhadap kebijakan pemerintah mengenai revisi Rancangan Undang-undang Tentara Negara Indonesia, dengan menerapkan metode IndoBERTweet. Selain itu, penelitian ini juga menganalisis bagaimana pengaruh normalisasi terhadap hasil akurasi dari metode IndoBERTweet. Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam bidang pengolahan data tekstual, khususnya dalam analisis sentimen metode IndoBERTweet, serta sebagai dasar pertimbangan bagi pembuat kebijakan di Indonesia agar lebih responsif terhadap aspirasi masyarakat. Penulis menyadari bahwa skripsi ini tidak akan dapat terselesaikan tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Allah SWT, Puji dan Syukur penulis senantiasa panjatkan kepada Allah SWT karena berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan Tugas Akhir ini.
2. Keluarga penulis, Terimakasih Umi, Bapak dan Teh Imas atas bimbingan, didikan, support dan doa yang telah diberikan kepada penulis.
3. Bapak Ir. Iwan Adhicandra, S.T., M.Sc., Ph.D., SMIEEE, selaku Ketua Program Studi Teknik Informatika dan juga selaku dosen pembimbing tugas akhir, memberikan ilmu, bimbingan, saran dan dukungannya kepada penulis dalam penelitian ini.
4. Bapak Albert Arapenta Sembiring, S.T., M.T, selaku dosen Penguji 1, telah memberikan ilmu dan saran kepada penulis.

5. Bapak Prof. Dr. Hoga Saragih, S.T., M.T, selaku Pembimbing Akademik dan juga sebagai dosen Penguji 2.
6. Bapak Irwan Prasetya Gunawan, S.T., M.Eng., PhD, selaku dosen akademik saya, yang sangat membantu dalam proses pembelajaran saya.
7. Semua Bapak/Ibu dosen Program Studi Informatika yang telah memberikan banyak ilmu, pengetahuan, wawasan selama penulis kuliah di Universitas Bakrie.
8. Teman-teman dan rekan seperjuangan yang berada di kampus Universitas Bakrie terutama Tiara Kalbia Nur ataupun yang diluar kampus telah memberikan arahan dan juga semangat kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat dan kontribusi positif bagi perkembangan ilmu pengetahuan dan teknologi.

Akhir kata, penulis mengucapkan terima kasih dan berharap semoga Allah SWT senantiasa memberikan rahmat dan karunia-Nya kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini.

Jakarta, 02 September 2025



Isnaeni Zakiyah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di

bawah ini: Nama : Isnaeni Zakiyah
NIM : 1212001025
Program Studi : Informatika
Fakultas : Teknik Dan Ilmu Komputer
Jenis Tugas Akhir : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul:

“PENGARUH NORMALISASI TEKS TERHADAP ANALISIS SENTIMEN BERBASIS INDOBERTWEET PADA TWEET TENTANG RUU TNI”

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada Tanggal : 02 September 2025

Yang Menyatakan,



Isnaeni Zakiyah

PENGARUH NORMALISASI TEKS TERHADAP ANALISIS SENTIMEN BERBASIS INDOBERTWEET PADA TWEET TENTANG RUU TNI

Isnaeni Zakiyah

ABSTRAK

Penelitian ini menganalisis pengaruh normalisasi teks terhadap performa analisis sentimen pada media sosial X (Twitter) terkait revisi RUU TNI. Data penelitian berupa 8.603 tweet yang dikumpulkan melalui web crawling pada periode 29 Maret–4 April 2025. Proses *preprocessing* meliputi cleaning, case folding, dan normalisasi, kemudian data dilabeli otomatis menjadi tiga kelas sentimen: positif, netral, dan negatif. Model IndoBERTweet digunakan dengan dua skenario, yaitu dataset normalisasi dan non-normalisasi. Evaluasi dilakukan menggunakan akurasi, *presisi*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa normalisasi teks tidak selalu meningkatkan kinerja model secara signifikan. Pada beberapa kondisi, data tanpa normalisasi memberikan hasil yang sebanding bahkan lebih baik. Sehingga, pengaruh normalisasi bersifat kontekstual dan bergantung pada karakteristik data serta model yang digunakan. Temuan ini berkontribusi pada pengembangan analisis sentimen berbahasa Indonesia, khususnya dalam isu sosial-politik, serta membuka peluang penelitian lanjutan dengan pendekatan ensemble learning dan domain-adaptive pretraining.

Kata kunci: Analisis Sentimen, IndoBERTweet, Normalisasi Teks, Media Sosial, RUU TNI

THE IMPACT OF TEXT NORMALIZATION ON SENTIMENT ANALYSIS OF TWEETS REGARDING THE TNI BILL USING INDOBERTTWEET

Isnaeni Zakiyah

ABSTRACT

This study examines the effect of text normalization on sentiment analysis performance of tweets regarding the revision of the TNI Bill. The dataset consists of 8,603 tweets collected through web crawling between March 29 and April 4, 2025. Preprocessing steps included cleaning, case folding, and normalization, followed by automatic labeling into three sentiment classes: positive, neutral, and negative. The IndoBERTTweet model was tested in two scenarios, normalized and non-normalized datasets. Performance was evaluated using accuracy, precision, recall, and F1-score. The results reveal that text normalization does not consistently improve model performance. In certain cases, non-normalized data achieved comparable or even better outcomes. The study concludes that the impact of normalization is highly contextual, depending on data characteristics and the chosen model. These findings contribute to the development of Indonesian sentiment analysis in socio-political domains and suggest further research with ensemble learning and domain-adaptive pretraining.

Keywords: Sentiment Analysis, IndoBERTTweet, Text Normalization, Social Media, RUU TNI

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	i
HALAMAN PENGESAHAN	ii
KATA PENGANTAR	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian.....	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori.....	9
2.2.1 Twitter	10

2.2.2	Analisis Sentimen	11
2.2.3	<i>Natural Language Processing</i> (NLP).....	11
2.2.4	<i>Machine Learning</i> (ML)	12
2.2.5	<i>Deep Learning</i>	12
2.2.6	IndoBERTweet	12
2.2.7	<i>Preprocessing Text</i>	13
2.2.8	Normalisasi Teks dan Pengaruhnya terhadap Kinerja Model ..	14
2.2.9	<i>Confusion Matrix</i>	15
2.2.10	Google Colaboratory	16
2.2.11	Python	17
2.2.12	Hugging Face.....	18
BAB III METODOLOGI PENELITIAN.....		19
3.1	Pendekatan Penelitian	19
3.2	Metodologi Penelitian	19
3.2.1	Metode Pengumpulan Data	20
A.	Data Sekunder.....	20
B.	Data Primer.....	20
3.2.2	Tahap Preprocessing Data	22
3.2.3	<i>Lebelling</i> Data.....	24
3.2.4	<i>Split Dataset</i>	24
3.2.5	Tokenisasi.....	24
3.2.6	Tahapan <i>Fine-Tuning</i>	25
3.2.7	Evaluasi (<i>Confusion Matrix</i>)	26
3.3	<i>Timeline</i> Penelitian	27

BAB IV HASIL DAN PEMBAHASAN	28
4.1 Deskripsi Data Penelitian	28
4.2 Hasil <i>Preprocessing</i>.....	29
4.2.1 <i>Cleaning Data</i>	29
4.2.1 <i>Case Folding</i>	31
4.2.2 Normalisasi	32
4.3 <i>Lebelling Data</i>	34
4.4 <i>Split Dataset</i>.....	37
4.5 Tokenisasi.....	39
4.6 Implementasi Fine-Tuning	41
4.7 Evaluasi Model (<i>Confusion Matrix</i>)	45
4.8 Hasil Perbedaan Matrix	53
4.8.1 Analisis Penyebab Perbedaan Performa	53
4.8.2 Implikasi Praktis	54
4.8.3 Signifikansi Temuan.....	55
BAB V KESIMPULAN DAN SARAN.....	56
5.1 Kesimpulan.....	56
5.2 Saran	57
Referensi	58

DAFTAR GAMBAR

Gambar 2.1 Report Hootsuite dan We Are Social	10
Gambar 2.2 Arsitektur IndoBERTweet.....	13
Gambar 2.3 Popularitas Bahasa Pemrograman, 2025.....	17
Gambar 3.1 Flowchart Penelitian.....	20
Gambar 3.2 Grafik Jumlah Cuitan Dataset RUU TNI	22
Gambar 4.1 Proses Hapus Dataset Duplikat	29
Gambar 4.2 Proses Cleaning Data dan Case Folding	30
Gambar 4.3 Proses Labelling	35
Gambar 4.4 Hasil Kelas Dataset Non-Normalisasi.....	36
Gambar 4.5 Hasil Kelas Dataset Normalisasi	36
Gambar 4.6 Grafik Kelas Analisis Sentimen	37
Gambar 4.7 Proses Split Dataset.....	38
Gambar 4.8 Grafik Hasil Split Dataset	38
Gambar 4.9 Proses Tokenisasi	39
Gambar 4.10 Proses Fine-Tuning	42
Gambar 4.11 Hasil Fine-Tuning Dataset Normalisasi	42
Gambar 4.12 Hasil Fine-Tuning Dataset Non-Normalisasi.....	43
Gambar 4.13 Grafik Perbandingan Hasil Training Loss.....	44
Gambar 4.14 Grafik Perbandingan Validation Accuracy dan F1-Macro per Epoch.....	44
Gambar 4.15 Perhitungan presisi, recall dan f-1 score	45
Gambar 4.16 Confusion Matrix Normalisasi	47
Gambar 4.17 Confusion Matrix Non-Normalisasi.....	49
Gambar 4.18 Grafik Perbandingan Confusion Matrix per Kelas.....	51
Gambar 4.19 Hasil Perbandingan Macro Average	52

DAFTAR TABEL

Tabel 2.1 Kajian Pustaka.....	6
Tabel 2.2 Confusion Matrix	16
Tabel 3 1 Case Folding	22
Tabel 3 2 Data Cleaning.....	23
Tabel 3 3 Hasil Normalisasi.....	23
Tabel 3 4 Konfigurasi Fine-Tuning.....	25
Tabel 3 5 Timeline Penelitian	27
Tabel 4.1 Hasil Cleaning Data dan Case Folding	31
Tabel 4.2 Proses Normalisasi Dataset.....	33
Tabel 4.3 Hasil Normalisasi.....	34
Tabel 4. 4 Hasil Labelling Dataset.....	35
Tabel 4. 5 Hasil Tokenisasi.....	40
Tabel 4. 6 Perbandingan Hasil Fine-Tuning	44