

**ANALISIS SENTIMEN PADA DATASET BERBAHASA INDONESIA
BERBASIS BERT**

TUGAS AKHIR

Oleh:

MUHAMMAD FACHRURRAZI

NIM.1910817310004



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2023

**ANALISIS SENTIMEN PADA DATASET BERBAHASA INDONESIA
BERBASIS BERT**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Sarjana Strata-1 Teknologi Informasi

Oleh:

MUHAMMAD FACHRURRAZI

NIM.1910817310004



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN, MEI 2023

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Muhammad Fachrurrazi
NIM : 1910817310004
Fakultas : Teknik
Program Studi : Teknologi Informasi
Judul Tugas Akhir : Analisis Sentimen pada Dataset Berbahasa
Indonesia Berbasis BERT
Pembimbing Utama : Ir. Yuslena Sari, S.Kom., M.Kom.

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, 12 Juni 2023



Muhammad Fachrurrazi

NIM. 1910817310004

LEMBAR PENGESAHAN
SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
Analisis Sentimen Pada Dataset Berbahasa Indonesia Berbasis BERT
oleh
Muhammad Fachrurrazi (1910817310004)

Telah dipertahankan di depan Tim Penguji pada 20 Juni 2023 dan
dinyatakan

L U L U S


Komite Penguji :

Ketua : Andreyan Rizky Baskara, S.Kom., M.Kom.
NIP 199307032019031011

Anggota 1 : Nurul Fathanah Mustamin, S.Pd., M.T.
NIP 199110252019032018

Anggota 2 : Andry Fajar Zulkarnain, S.ST., M.T
NIP 199007272019031018

Pembimbing : Dr. Ir. Yuslena Sari, S.Kom., M.Kom.
Utama NIP 198411202015042002



Banjarbaru, ..04..JUL..2023....
diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik
Fakultas Teknik ULM,



Dr. Wahyuni, S.T., M.T.
NIP 197401071998021001

Koordinator Program Studi
S-1 Teknologi Informasi,



Dr. Ir. Yuslena Sari, S.Kom., M.Kom.
NIP 198411202015042002

PERSETUJUAN TUGAS AKHIR

ANALISIS SENTIMEN PADA DATASET BERBAHASA INDONESIA
BERBASIS BERT

OLEH
MUHAMMAD FACHRURRAZI
NIM. 1910817310004

Telah terpenuhi dan terpenuhi semua persyaratan akademik, administrasi dan
telah disetujui untuk dipertahankan di hadapan dewan penguji

Banjarmasin, 9 Juni 2023

Pembimbing Utama,



Dr. Ir. Yusefena Sari S.Kom., M.Kom.

NIP. 198411202015042002

ABSTRAK

Banyak model analisis sentimen yang tersedia berasal dari penelitian-penelitian sebelumnya yang memiliki cakupan yang spesifik pada topik atau produk tertentu. Cakupan yang spesifik merujuk pada dataset yang dibuat untuk tujuan tertentu, seperti misalnya untuk mengenali sentimen pada produk makanan atau pada politikus tertentu. Dataset yang spesifik juga dapat menghasilkan bias pada hasil analisis sentimen yang dilakukan karena cakupan yang terbatas hanya pada topik atau produk tertentu. Dengan melakukan analisis sentimen pada dataset dengan tidak terfokus pada topik tertentu, maka dapat mengidentifikasi pola perilaku atau opini umum di kalangan masyarakat. *Natural Language Processing* dapat digunakan untuk melakukan *text classification*. BERT, MBERT, dan IndoBERT merupakan salah satu *pre-trained* model yang dapat melakukan proses NLP. Dataset yang digunakan berasal dari berbagai sumber platform media sosial dengan jumlah total 12.760 data. Model IndoBERT memberikan performa model yang terbaik menggunakan konfigurasi rasio pembagian data sebanyak 90:10, untuk nilai 90% data (11,484 data) digunakan untuk proses *training* dan validasi, serta 10% data (1.276 data) digunakan untuk proses *testing*, 5 *epoch* dan *batch size* 16 mendapatkan hasil terbaik dengan nilai tingkat akurasi sebesar 92,71%, f1-score sebesar 89,82%, recall sebesar 89,02%, dan presisi sebesar 90,86%. Performa yang didapatkan lebih baik dibandingkan dengan model BERT dan MBERT dalam berbagai konfigurasi. model IndoBERT mampu mengklasifikasikan dengan benar sebanyak 1183 data dari total 1276 data uji.

Kata kunci: Analisis Sentimen, BERT, IndoBERT, MBERT, NLP

ABSTRACT

Many sentiment analysis models have been derived from previous research with specific scopes on particular topics or products. Specific scope refers to datasets created for specific purposes, such as identifying sentiments of food products or specific politicians. Specific datasets can introduce bias in sentiment analysis results, owing to the limited coverage of specific topics or products. By conducting sentiment analysis on datasets that do not focus on specific topics, patterns of behavior or general opinions among the public can be identified. Natural Language Processing can also be used for text classification. BERT, MBERT, and IndoBERT are pre-trained models that can perform NLP processes. The dataset used was obtained from various social media platforms, with 12,760 data points. The IndoBERT model provided the best performance using a data split configuration of 90:10, where 90% of the data (11,484 data points) were used for training and validation and 10% of the data (1,276 data points) were used for testing. With 5 epochs and a batch size of 16, the model achieved the best results, with an accuracy rate of 92.71%, an F1-score of 89.82%, a recall of 89.02%, and a precision of 90.86%. The performance obtained was better than those of the BERT and MBERT models in various configurations. The IndoBERT model correctly classified 1,183 of 1,276 test data points.

Keywords: IndoBERT, BERT, MBERT, NLP, Sentiment Analysis

LEMBAR PERSEMBAHAN

Penulis mempersembahkan Tugas Akhir ini kepada:

1. Almarhum Ayah terkasih yang sepanjang hidupnya sudah memberikan dukungan tanpa henti telah menginspirasi dan mendorong penulis dalam menyelesaikan studi. Meskipun tidak lagi berada disini, kenangan dan arahan Ayah tetap menjadi pendorong utama keberhasilan Tugas Akhir ini.
2. Kepada Ibu, Kakak serta keluarga, terima kasih atas cinta, dukungan, doa yang tak pernah henti untuk penulis dalam keberlangsungan penyelesaian Tugas Akhir ini yang penulis terima.
3. Ibu Ir. Yuslena Sari, S.Kom., M.Kom. selaku Koordinator Tugas Akhir dan sekaligus Dosen Pembimbing Utama yang telah memberikan banyak motivasi, kesempatan, bimbingan, arahan, dan ilmu baik itu dari penyusunan Laporan Tugas Akhir, akademik, moral dan etika yang berkaitan dengan bidang dan keilmuan penulis terima.
4. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dalam penyelesaian Laporan Tugas Akhir.
5. Teman seperjuangan dari masa perkuliahan sampai Tugas Akhir, Muhammad Afrizal Miqdad, Muhammad Nur Abdi, Irvan Aulia Luthfi, Muhammad Faidhorrahman, Muhammad Ardhy Satrio Jati, Muhammad Arras, Digdo Aji Asrowi, Muhammad Miftachul Falah, serta seluruh teman-teman Angkatan 2019 Program Studi Teknologi Informasi.
6. Terakhir namun tak kalah penting, terima kasih kepada diri sendiri atas kepercayaan diri yang telah dibangun, ketekunan yang tak pernah pudar, dan keberanian untuk menghadapi tantangan yang muncul. Serta atas kerja keras, dedikasi, dan semangat yang ditanamkan dalam diri sendiri dalam mengerjakan Tugas Akhir ini hingga akhirnya bisa menyelesaikan Tugas Akhir ini.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Allah Subhanahu wa ta'ala yang telah melimpahkan rahmat dan hidayah-Nya. Sehingga penulis dapat menyelesaikan Tugas Akhir dengan Judul “Analisis Sentimen pada Dataset Berbahasa Indonesia Berbasis BERT”.

Dalam penyusunan laporan ini, penulis menyadari sepenuhnya bahwa laporan ini tidak akan terwujud tanpa adanya bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Almarhum Ayah terkasih yang sepanjang hidupnya sudah memberikan dukungan tanpa henti telah menginspirasi dan mendorong penulis dalam menyelesaikan studi. Meskipun tidak lagi berada disini, kenangan dan arahan Ayah tetap menjadi pendorong utama keberhasilan Tugas Akhir ini.
2. Kepada Ibu, Kakak serta keluarga, terima kasih atas cinta, dukungan, doa yang tak pernah henti untuk penulis dalam keberlangsungan penyelesaian Tugas Akhir ini yang penulis terima.
3. Ibu Ir. Yuslena Sari, S.Kom., M.Kom. selaku Koordinator Tugas Akhir dan sekaligus Dosen Pembimbing Utama yang telah memberikan banyak motivasi, kesempatan, bimbingan, arahan, dan ilmu baik itu dari penyusunan laporan Tugas Akhir, akademik, moral dan etika yang berkaitan dengan bidang dan keilmuan penulis terima.
4. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dalam penyelesaian laporan Tugas Akhir.
5. Teman seperjuangan dari masa perkuliahan sampai Tugas Akhir, Muhammad Afrizal Miqdad, Muhammad Nur Abdi, Irvan Aulia Luthfi, Muhammad Faidhorrahman, Muhammad Ardhy Satrio Jati, Muhammad Arras, Digdo Aji Asrowi, Muhammad Miftachul Falah, serta seluruh teman-teman Angkatan 2019 Program Studi Teknologi Informasi.
6. Terakhir namun tak kalah penting, terima kasih kepada diri sendiri atas kepercayaan diri yang telah dibangun, ketekunan yang tak pernah pudar, dan keberanian untuk menghadapi tantangan yang muncul. Serta atas kerja keras, dedikasi, dan semangat yang ditanamkan dalam diri sendiri dalam

mengerjakan Tugas Akhir ini hingga akhirnya bisa menyelesaikan Tugas Akhir ini.

Akhir kata, penulis menyampaikan terima kasih kepada seluruh pihak yang ikut serta membantu dalam penyelesaian laporan Tugas Akhir ini. Penyusunan laporan Tugas Akhir ini disusun dengan sebaik-baiknya namun masih terdapat kekurangan di dalam penyusunan laporan Tugas Akhir ini. Oleh karena itu, saran dan kritik yang sifatnya membangun dari semua pihak sangat diharapkan, tidak lupa harapan penulis agar laporan ini dapat bermanfaat bagi pembaca

Banjarmasin, 12 Juni 2023

Penulis

Muhammad Fachrurrazi

DAFTAR ISI

HALAMAN SAMPUL LUAR.....	i
HALAMAN SAMPUL DALAM.....	ii
LEMBAR PERNYATAAN	iii
LEMBAR PENGESAHAN.....	iv
LEMBAR PERSETUJUAN.....	v
ABSTRAK	vi
ABSTRACT	vii
LEMBAR PERSEMBAHAN.....	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xviii
DAFTAR RIWAYAT HIDUP.....	xix
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1 Landasan Teori	6
2.1.1 Analisis Sentimen.....	6
2.1.2 <i>Natural Language Processing</i>	6
2.1.3 <i>Text Pre-processing</i>	6
2.1.4 <i>Transformers</i>	8
2.1.5 <i>Bidirectional Encoder Representations from Transformers (BERT)</i>	10
2.1.6 <i>Multilingual BERT (MBERT)</i>	11
2.1.7 <i>Indonesian based on the Bidirectional Encoder Representation from Transformers (IndoBERT)</i>	11
2.1.8 <i>Confusion Matrix</i>	12
2.2 Penelitian Terkait.....	13

2.2.1	<i>Sentiment Analysis of News Headlines Using Simple Transformers</i>	13
2.2.2	<i>Sentiment Analysis towards Cryptocurrency and NFT in Bahasa Indonesia for Twitter Large Amount Data Using BERT</i>	14
2.2.3	<i>Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT</i>	14
2.2.4	<i>Improving IndoBERT for Sentiment Analysis on Indonesian Stock Trader Slang Language</i>	15
2.2.5	<i>Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method</i>	16
2.2.6	<i>Sentiment Analysis of Cooking Oil using Bidirectional Encoder Representations from Transformers</i>	16
2.3	Kerangka Pemikiran	20
2.3.1	<i>Indicators</i>	20
2.3.2	<i>Proposed Method</i>	20
2.3.3	<i>Objectives</i>	20
2.3.4	<i>Measurement</i>	21
BAB III METODOLOGI PENELITIAN		22
3.1	Alat dan Bahan	22
3.1.1	Alat Penelitian	22
3.1.2	Bahan Penelitian	22
3.2	Prosedur Penelitian	22
3.2.1	Identifikasi Masalah	23
3.2.2	Studi Literatur	23
3.2.3	Pengambilan dataset	23
3.2.4	Melakukan Eksperimen	24
3.2.5	Evaluasi sistem	28
3.2.6	Implementasi Sistem	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Pengambilan Dataset	31
4.2	<i>Text Pre-processing</i>	31
4.2.1	<i>Case Folding</i>	32
4.2.2	<i>Filtering</i>	32
4.2.3	<i>Tokenizing</i>	32
4.2.4	<i>Normalization</i>	32
4.3	Implementasi Model	33

4.4	Eksperimen Model.....	36
4.4.1	Eksperimen Model BERT.....	37
4.4.2	Eksperimen Model MBERT.....	44
4.4.3	Eksperimen Model IndoBERT.....	51
4.5	Pengujian.....	58
4.1	Pengujian Pertama.....	58
4.2	Pengujian Kedua.....	58
4.3	Pengujian Ketiga.....	59
4.4	Pengujian Keempat.....	60
4.6	Evaluasi Model Terbaik.....	61
4.7	Implementasi Model IndoBERT pada Sistem.....	65
BAB V KESIMPULAN DAN SARAN.....		68
5.1.	Kesimpulan.....	68
5.2.	Saran.....	68
DAFTAR PUSTAKA.....		70
LAMPIRAN.....		75

DAFTAR TABEL

Tabel 2. 1 Perbedaan BERT, mBERT, dan IndoBERT	12
Tabel 2. 2 Confusion Matrix	12
Tabel 2. 3 Statistical Measurement IndoBERT	15
Tabel 2. 4 Ringkasan Penelitian Terkait	18
Tabel 3. 1 Alat Penunjang Penelitian.....	22
Tabel 3. 2 Bahan Penelitian.....	22
Tabel 3. 3 Contoh Data	24
Tabel 3. 4 Proses Case Folding	26
Tabel 3. 5 Proses Punctuation Removal.....	27
Tabel 3. 6 Proses Normalization.....	27
Tabel 3. 7 Proses Tokenizing	27
Tabel 4. 1 Hasil Case Folding	32
Tabel 4. 2 Hasil Filtering	32
Tabel 4. 3 Hasil Tokenizing	32
Tabel 4. 4 Hasil Normalization	33
Tabel 4. 5 Konfigurasi Parameter Eksperimen Pertama Model BERT	37
Tabel 4. 6 Hasil Eksperimen Pertama Model BERT	38
Tabel 4. 7 Konfigurasi Parameter Eksperimen Kedua Model BERT	39
Tabel 4. 8 Hasil Eksperimen Kedua Model BERT.....	39
Tabel 4. 9 Konfigurasi Parameter Eksperimen Ketiga Model BERT.....	40
Tabel 4. 10 Hasil Eksperimen Ketiga Model BERT	41
Tabel 4. 11 Konfigurasi Parameter Eksperimen Keempat Model BERT	42
Tabel 4. 12 Hasil Eksperimen Keempat Model BERT.....	43
Tabel 4. 13 Konfigurasi Parameter Eksperimen Pertama Model MBERT	44
Tabel 4. 14 Hasil Eksperimen Pertama Model MBERT.....	44
Tabel 4. 15 Konfigurasi Parameter Eksperimen Kedua Model MBERT	45
Tabel 4. 16 Hasil Eksperimen Kedua Model MBERT	46
Tabel 4. 17 Konfigurasi Parameter Eksperimen Ketiga Model MBERT	47
Tabel 4. 18 Hasil Eksperimen Ketiga Model MBERT	48
Tabel 4. 19 Konfigurasi Parameter Eksperimen Keempat Model MBERT.....	49

Tabel 4. 20 Hasil Eksperimen Keempat Model MBERT	50
Tabel 4. 21 Konfigurasi Parameter Eksperimen Pertama Model IndoBERT	51
Tabel 4. 22 Hasil Eksperimen Pertama Model IndoBERT	51
Tabel 4. 23 Konfigurasi Parameter Eksperimen Kedua Model IndoBERT	52
Tabel 4. 24 Hasil Eksperimen Kedua Model IndoBERT	53
Tabel 4. 25 Konfigurasi Parameter Eksperimen Ketiga Model IndoBERT	54
Tabel 4. 26 Hasil Eksperimen Ketiga Model IndoBERT	55
Tabel 4. 27 Konfigurasi Parameter Eksperimen Keempat Model IndoBERT	56
Tabel 4. 28 Hasil Eksperimen Keempat Model IndoBERT	57
Tabel 4. 29 Hasil Pengujian Pertama	58
Tabel 4. 30 Hasil Pengujian Kedua	58
Tabel 4. 31 Hasil Pengujian Ketiga	59
Tabel 4. 32 Hasil Pengujian Keempat	60
Tabel 4. 33 Konfigurasi Model IndoBERT	61
Tabel 4. 34 Perbandingan Performa Terbaik Setiap Pengujian	61
Tabel 4. 35 Hasil Pengujian IndoBERT menggunakan Confusion Matrix	62
Tabel 4. 36 Hasil Klasifikasi Data	62
Tabel 4. 37 Contoh Data yang Berhasil Diklasifikasi dengan Benar	63
Tabel 4. 38 Contoh Data yang Berhasil Diklasifikasi dengan Salah	63

DAFTAR GAMBAR

Gambar 2. 1 Model Arsitektur Transformer [23]	8
Gambar 2. 2 Arsitektur Encoder Decoder pada Transformer [24]	9
Gambar 2. 3 BERT Architecture [14]	10
Gambar 2. 4 Kerangka Pemikiran	20
Gambar 3. 1 Prosedur Penelitian	23
Gambar 3. 2 Tahapan proses implementasi model	25
Gambar 3. 3 Skema Pengujian	29
Gambar 3. 4 Rancangan implementasi sistem.....	30
Gambar 4. 1 Grafik nilai Accuracy dan Loss model BERT 3 epoch dan batch size 16	38
Gambar 4. 2 Grafik nilai Accuracy dan Loss model BERT 3 epoch dan batch size 32	38
Gambar 4. 3 Grafik nilai Accuracy dan Loss model BERT 5 epoch dan batch size 16	40
Gambar 4. 4 Grafik nilai Accuracy dan Loss model BERT 5 epoch dan batch size 32	40
Gambar 4. 5 Grafik nilai Accuracy dan Loss model BERT 3 epoch dan batch size 16	41
Gambar 4. 6 Grafik nilai Accuracy dan Loss model BERT 3 epoch dan batch size 32	42
Gambar 4. 7 Grafik nilai Accuracy dan Loss model BERT 5 epoch dan batch size 16	43
Gambar 4. 8 Grafik nilai Accuracy dan Loss model BERT 5 epoch dan batch size 32	43
Gambar 4. 9 Grafik nilai Accuracy dan Loss model MBERT 3 epoch dan batch size 16	45
Gambar 4. 10 Grafik nilai Accuracy dan Loss model MBERT 3 epoch dan batch size 32	45
Gambar 4. 11 Grafik nilai Accuracy dan Loss model MBERT 5 epoch dan batch size 16	47

Gambar 4. 12 Grafik nilai Accuracy dan Loss model MBERT 5 epoch dan batch size 32	47
Gambar 4. 13 Grafik nilai Accuracy dan Loss model MBERT 3 epoch dan batch size 16	48
Gambar 4. 14 Grafik nilai Accuracy dan Loss model MBERT 3 epoch dan batch size 32	49
Gambar 4. 15 Grafik nilai Accuracy dan Loss model IndoBERT 3 epoch dan batch size 16	52
Gambar 4. 16 Grafik nilai Accuracy dan Loss model IndoBERT 3 epoch dan batch size 32	52
Gambar 4. 17 Grafik nilai Accuracy dan Loss model IndoBERT 5 epoch dan batch size 16	54
Gambar 4. 18 Grafik nilai Accuracy dan Loss model IndoBERT 5 epoch dan batch size 32	54
Gambar 4. 19 Grafik nilai Accuracy dan Loss model IndoBERT 3 epoch dan batch size 16	55
Gambar 4. 20 Grafik nilai Accuracy dan Loss model IndoBERT 3 epoch dan batch size 32	56
Gambar 4. 21 Grafik nilai Accuracy dan Loss model IndoBERT 5 epoch dan batch size 16	57
Gambar 4. 22 Grafik nilai Accuracy dan Loss model IndoBERT 5 epoch dan batch size 32	57
Gambar 4. 23 Confusion Matrix IndoBERT	62
Gambar 4. 24 Antarmuka sistem analisis sentimen berbahasa Indonesia.....	65
Gambar 4. 25 Contoh Masukan Kalimat Sentimen Negatif.....	66
Gambar 4. 26 Contoh Masukan Kalimat Sentimen Positif	66
Gambar 4. 27 Contoh Masukan Kalimat Sentimen Netral	67