

**IMPLEMENTASI INDOBERT UNTUK ANALISIS
SENTIMEN KOMENTAR YOUTUBE TERHADAP PIDATO
PRESIDEN DI SIDANG PARIPURNA DPR RI**

Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



**Widya Rahmawati Putri
1313622008**

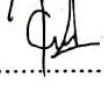
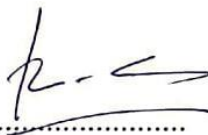
**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA**

2026

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

IMPLEMENTASI INDOBERT UNTUK ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP PIDATO PRESIDEN DI SIDANG PARIPURNA DPR RI

Nama : Widya Rahmawati Putri
No. Registrasi : 1313622008

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Dr. Hadi Nasbey, S.Pd., M.Si.</u> NIP. 197909162005011004		20-08-2026
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Meiliasari, S.Pd., M.Sc.</u> NIP. 197905042009122002		20-08-2026
Ketua Penguji	: <u>Drs. Mulyono, M.Kom.</u> NIP. 196605171994031003		04-08-2026
Sekretaris	: <u>Carli Apriansyah Hutagalung, M.Kom.</u> NIP. 199604102025061008		05-08-2026
Anggota			
Pembimbing I	: <u>Dr. Ria Arafiah, M. Si.</u> NIP. 197511212005012004		07-08-2026
Pembimbing II	: <u>Dinar Munggaran Akhmad, S.Kom., M.Kom.</u> NIP. 199210032025062003		06-08-2026
Penguji Ahli	: <u>Ari Hendarno, S.Pd, M.Kom.</u> NIP. 19881102202031002		05-08-2026

Dinyatakan lulus ujian skripsi tanggal: 04 Agustus 2026

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sebenar-benarnya bahwa skripsi berjudul **“Implementasi IndoBERT untuk Analisis Sentimen Komentar YouTube terhadap Pidato Presiden di Sidang Paripurna DPR RI”**, yang disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari penulis lain yang telah dipublikasikan yang disebutkan dalam teks skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Jika di kemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 15 Juli 2026
Yang membuat pernyataan



Widya Rahmawati Putri

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, serta penyertaan-Nya, sehingga skripsi yang berjudul *“Implementasi IndoBERT untuk Analisis Sentimen Komentar YouTube terhadap Pidato Presiden di Sidang Paripurna DPR RI”* ini dapat diselesaikan dengan baik. Keberhasilan dalam penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan berbagai pihak yang dengan tulus dan ikhlas telah memberikan masukan yang bermanfaat kepada penulis. Pada kesempatan ini, penulis mengucapkan terima kasih kepada:

1. Dr. Ria Arafiah, M.Si, selaku dosen pembimbing pertama dan selaku Koordinator Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta yang telah memberikan bimbingan, arahan, serta dukungan dengan penuh kesabaran selama proses penyusunan skripsi ini.
2. Dinar Munggaran Akhmad, S.Kom., M.Kom, selaku dosen pembimbing kedua yang telah memberikan bimbingan, koreksi, saran, serta arahan yang sangat berarti dalam penyusunan skripsi ini.
3. Kedua orang tua dan keluarga penulis yang telah memberikan doa dan menyemangati penulis agar segera lulus.
4. Kepada semua teman-teman penulis seperti Devi, Jore, Wulan, Evi, Rista, Fidya dan teman dedomena lainnya yang telah membantu penulis dan memberikan dukungan semangat untuk penelitian penulis.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari sempurna. Oleh karena itu, penulis memohon maaf apabila terdapat kesalahan maupun kekeliruan dalam penulisan skripsi ini.

Jakarta, 15 Juli 2026



Widya Rahmawati Putri

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA
ILMIAH UNTUK KEPENTINGAN AKADEMIS**



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Widya Rahmawati Putri
NIM : 1313622008
Fakultas/Prodi : FMIPA/ Ilmu Komputer
Alamat email : wrahmawatiputri@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain

yang berjudul :

Implementasi IndoBERT untuk Analisis Sentimen Komentar YouTube terhadap Pidato Presiden di Sidang Paripurna DPR RI

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 21 Agustus 2026

(Widya Rahmawati Putri)

ABSTRAK

Widya Rahmawati Putri. Implementasi IndoBERT untuk Analisis Sentimen Komentar YouTube terhadap Pidato Presiden di Sidang Paripurna DPR RI. Skripsi, Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Juli 2026.

Perubahan kondisi ekonomi Indonesia yang dipengaruhi oleh pelemahan nilai tukar rupiah, ketidakpastian ekonomi global, serta tantangan dalam kebijakan fiskal memunculkan beragam respons masyarakat. Salah satu respons tersebut muncul setelah Presiden Prabowo Subianto menyampaikan pidato pada Sidang Paripurna DPR RI tanggal 20 Mei 2026 mengenai Kerangka Ekonomi Makro dan Pokok-Pokok Kebijakan Fiskal RAPBN Tahun 2027. Berbagai tanggapan masyarakat disampaikan melalui kolom komentar pada platform YouTube sehingga menjadi sumber data yang dapat dimanfaatkan untuk mengetahui kecenderungan opini publik terhadap kebijakan pemerintah. Penelitian ini bertujuan mengimplementasikan model IndoBERT (*Indonesia Bidirectional Encoder Representations from Transformers*) untuk mengklasifikasikan sentimen komentar YouTube ke dalam tiga kelas, yaitu positif, negatif, dan netral. Data penelitian dikumpulkan menggunakan YouTube Data API v3 dari delapan kanal berita nasional, yaitu CNN Indonesia, CNN, DPR RI, Kompas TV, KompasTV Madiun, Kumparan, Metro TV, dan Sekretariat Presiden, yang mengunggah delapan video pidato Presiden pada Sidang Paripurna DPR RI. Proses pengumpulan data menghasilkan 3.071 komentar. Selanjutnya, melakukan pelabelan data, kemudian dilakukan tahapan *pre-processing* yang meliputi *case folding*, *cleaning*, *normalisasi* kata tidak baku, *tokenisasi*, dan penghapusan *stopword*. Data yang telah diproses kemudian digunakan pada tahap *fine-tuning* model IndoBERT untuk melakukan klasifikasi sentimen. Kinerja model dievaluasi menggunakan confusion matrix dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, kemudian hasil klasifikasi dianalisis untuk mengetahui distribusi sentimen masyarakat terhadap pidato Presiden. Hasil evaluasi menunjukkan bahwa model IndoBERT memperoleh *accuracy* sebesar 88,72%, *precision* 85,13%, *recall* 81,78%, dan *F1-score* 83,02%. Distribusi hasil klasifikasi menunjukkan bahwa sentimen positif mendominasi dibandingkan sentimen negatif dan netral, yang menunjukkan bahwa komentar pada dataset penelitian didominasi oleh sentimen positif terhadap kebijakan yang dipaparkan dalam pidato Presiden. Hasil penelitian menunjukkan bahwa model IndoBERT menunjukkan performa klasifikasi yang baik pada dataset penelitian dalam melakukan klasifikasi sentimen pada teks berbahasa Indonesia serta berpotensi dimanfaatkan untuk menganalisis kecenderungan opini publik terhadap isu kebijakan pemerintah.

Kata kunci: Analisis sentimen, IndoBERT, YouTube, Pidato Presiden.

ABSTRACT

Widya Rahmawati Putri. Implementation of IndoBERT for Sentiment Analysis of YouTube Comments on the President's Speech at the Plenary Session of the House of Representatives of the Republic of Indonesia (DPR RI). Thesis, Computer Science Study Program, Faculty of Mathematics and Natural Sciences, State University of Jakarta. July 2026.

Shifts in Indonesia's economic conditions driven by a weakening rupiah, global economic pressures, and fiscal policy challenges have elicited diverse public reactions. One such response emerged following President Prabowo Subianto's address at the House of Representatives (DPR RI) Plenary Session on May 20, 2026, regarding the Macroeconomic Framework and Fiscal Policy Principles for the 2027 State Budget (RAPBN). Public feedback was expressed via YouTube comment sections, serving as a data source to gauge public opinion trends regarding government policy. This study aims to implement the IndoBERT (Indonesia BiDirectional Encoder Representations from Transformers) model to classify YouTube comments into three sentiment categories: positive, negative, and neutral. Data were collected using the YouTube Data API v3 from eight national news channels CNN Indonesia, CNN, DPR RI, Kompas TV, KompasTV Madiun, Kumparan, Metro TV, and the Presidential Secretariat which uploaded videos of the President's speech at the DPR RI Plenary Session. The data collection process yielded 3,071 comments. Subsequently, the data underwent labeling and a pre-processing stage comprising case folding, cleaning, non-standard word normalization, tokenization, and stopword removal. The processed data were then used to fine-tune the IndoBERT model for sentiment classification. Model performance was evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics, and the classification results were analyzed to determine the distribution of public sentiment regarding the President's speech. Evaluation results indicate that the IndoBERT model achieved an accuracy of 88.72%, precision of 85.13%, recall of 81.78%, and an F1-score of 83.02%. The distribution of classification results shows that positive sentiment predominated over negative and neutral sentiments, indicating that opinions within the research dataset were largely positive regarding the policies outlined in the President's speech. Research results indicate that the IndoBERT model demonstrates good classification performance on the study dataset for sentiment classification of Indonesian text and has the potential to be utilized for analyzing public opinion trends regarding government policy issues.

Keywords: Sentiment analysis, IndoBERT, YouTube, Presidential speech.

DAFTAR ISI

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI.....	ii
LEMBAR PERNYATAAN.....	iii
KATA PENGANTAR	iv
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II KAJIAN PUSTAKA.....	5
2.1 Analisis Sentimen.....	5
2.2 Pidato Presiden di Sidang Paripurna DPR	6
2.2.1 Komentar Pidato Presiden	6
2.3 <i>Text Mining</i>	7
2.4 <i>Natural Language Processing (NLP)</i>	8
2.5 <i>Deep Learning</i>	9
2.6 <i>Transformer</i>	10
2.7 <i>Indonesia Bidirectional Encoder Representations from Transformers (IndoBERT)</i>	15

2.8	<i>Fleiss Kappa</i>	21
2.9	<i>Confusion Matrix</i>	22
2.10	Penelitian Terdahulu.....	24
BAB III METODE PENELITIAN.....		28
3.1	Tahapan Penelitian.....	28
3.1.1	Pengumpulan Data.....	28
3.1.2	Pelabelan Data.....	33
3.1.3	<i>Pre-Processing Data</i>	34
3.1.4	Klasifikasi Model IndoBERT.....	36
3.1.5	Evaluasi dan Analisis Data.....	42
3.2	Jadwal Penelitian.....	42
BAB IV HASIL DAN PEMBAHASAN.....		43
4.1	Pengumpulan Data.....	43
4.2	Pelabelan Data.....	44
4.3	<i>Pre- processing Data</i>	50
4.3.1	<i>Case Folding</i>	51
4.3.2	<i>Cleaning</i>	52
4.3.3	<i>Normalisasi Slang</i>	54
4.3.4	Tokenisasi.....	56
4.3.5	Penghapusan Stopword.....	57
4.4	Klasifikasi IndoBERT.....	60
4.4.1	Pembagian Data.....	61
4.4.2	Penyeimbangan Data (<i>Random Oversampling</i>).....	61
4.4.3	<i>WordPiece Tokenizer</i>	62
4.4.4	Model <i>Fine-Tuning</i> IndoBERT.....	63
4.4.5	Hasil Klasifikasi Model IndoBERT.....	66

4.5 Evaluasi Model.....	67
4.6 Analisis Distribusi Sentimen Komentar YouTube terhadap Pidato Presiden	70
4.7 Analisis Kesalahan Klasifikasi	72
4.7.1 Analisis Kemampuan IndoBERT dalam Mengenali Sarkasme	73
4.7.2 Analisis Karakteristik Linguistik Berdasarkan Kesalahan Klasifikasi.....	74
4.7.3 Solusi Penyebab Kesalahan Klasifikasi pada IndoBERT	76
4.8 Visualisasi WordCloud	77
BAB V KESIMPULAN DAN SARAN.....	79
5.1 Kesimpulan.....	79
5.2 Saran	80
DAFTAR PUSTAKA.....	81
LAMPIRAN	85
RIWAYAT HIDUP.....	107



DAFTAR GAMBAR

Gambar 2. 1 Representasi jaringan saraf	9
Gambar 2. 2 Arsitektur Transformer	10
Gambar 2. 3 Lapisan Encoder	11
Gambar 2. 4 Lapisan Decoder	12
Gambar 2. 5 Scaled Dot Product Attention	13
Gambar 2. 6 multi head self-attention	13
Gambar 2. 7 Arsitektur BERT	16
Gambar 2. 8 Architecture of the IndoBERT model	17
Gambar 2. 9 Masked Language Model	18
Gambar 2. 10 Next Sentence Prediction.....	19
Gambar 2. 11 Fine-Tuning BERT.....	19
Gambar 2. 12 Confusion Matrix.....	23
Gambar 3. 1 Alur Penelitian	28
Gambar 3. 2 Alur Pre-Processing	34
Gambar 3. 3 Alur Klasifikasi.....	36
Gambar 3. 4 Input data bersih	37
Gambar 3. 5 Proses Tokenisasi IndoBERT.....	38
Gambar 3. 6 Proses konversi ke input ID.....	38
Gambar 3. 7 Proses attention MASK	39
Gambar 3. 8 Proses Input embedding BERT.....	40
Gambar 3. 9 Proses representasi Token [CLS].....	40
Gambar 3. 10 Prapelatihan dan Fine-tuning IndoBERT	41
Gambar 4. 1. Hasil Crawling data komentar YouTube.....	44
Gambar 4. 2 Upload data mentah di Google Colab.....	50
Gambar 4. 3 Fungsi case folding	51
Gambar 4. 4 Fungsi cleaning	53
Gambar 4. 5 Fungsi Normalisasi slang.....	55
Gambar 4. 6 Fungsi tokenisasi	56
Gambar 4. 7 Fungsi penghapusan stopword.....	58
Gambar 4. 8 Fungsi keseluruhan teks bersih.....	59

Gambar 4. 9 Perbandingan kelas data latih Sebelum dan Sesudah Oversampling	62
Gambar 4. 10 Hasil proses perhitungan tokenisasi.....	63
Gambar 4. 11 Kurva Training dan Validation Loss	66
Gambar 4. 12 Confusion Matrix Data Uji	68
Gambar 4. 13 Distribusi Sentimen Publik terhadap Pidato Presiden	70
Gambar 4. 14. WordCloud untuk Kelas Negatif, Netral, dan Positif	77



DAFTAR TABEL

Tabel 2. 1 Korpus IndoBERT	17
Tabel 2. 2 Interpretasi Fleiss' Kappa	22
Tabel 2.3 Penelitian Terdahulu	24
Tabel 3.1 Contoh Komentar berita di Youtube	29
Tabel 3. 2 Latar belakang anotator	33
Tabel 3.3 Case Folding	34
Tabel 3.4 Cleaning	35
Tabel 3.5 Normalisasi Slang	35
Tabel 3.6 Tokenisasi	35
Tabel 3.7 Penghapusan Stopword	36
Tabel 3.8 Jadwal Penelitian	42
Tabel 4. 1 Jumlah komentar per kanal YouTube	43
Tabel 4. 2 Contoh pelabelan data	45
Tabel 4. 3 Sampel label data kappa value	46
Tabel 4. 4 Contoh data mentah	50
Tabel 4. 5 Hasil proses case folding	51
Tabel 4. 6 Hasil proses cleaning	53
Tabel 4. 7 Hasil proses normalisasi slang	55
Tabel 4. 8 Hasil proses tokenisasi	56
Tabel 4. 9 Hasil proses penghapusan stopword	58
Tabel 4. 10 Hasil Komentar Teks Bersih	60
Tabel 4. 11 Hasil Pembagian Data Latih, Validasi, dan Uji	61
Tabel 4. 12 Perbandingan hyperparameter fine-tuning	63
Tabel 4. 13 Hyperparameter yang digunakan Fine-Tuning	64
Tabel 4. 14 Proses per epoch selama training dan validasi	65
Tabel 4. 15 Hasil dari Klasifikasi	66
Tabel 4. 16 Classification Report per Kelas Sentimen pada Data Uji	67
Tabel 4. 17 Hasil Evaluasi Akhir pada Data Uji	68
Tabel 4. 18 Contoh Prediksi Model pada Komentar Baru	70
Tabel 4. 19 Hasil Kesalahan Klasifikasi	72

Tabel 4. 20. Contoh komentar Sarkasme yang terdeteksi IndoBERT	73
Tabel 4. 21. Contoh komentar Sarkasme yang tidak terdeteksi IndoBERT	73
Tabel 4. 22 Distribusi Label Sentimen pada Komentar Berfrasa “omon-omon”	75
Tabel 4. 23 Data komentar sarkasme “omon omon”	75



DAFTAR LAMPIRAN

Lampiran 1. Surat Tugas Anotator 1.....	85
Lampiran 2. Surat Tugas Anotator 2.....	86
Lampiran 3. Surat Tugas Anotator 3.....	87
Lampiran 4. Code Crawling Data Komentar Youtube.....	88
Lampiran 5. Code Labeling Data Komentar	88
Lampiran 6. Code Fleiss' Kappa	89
Lampiran 7. Code Preprocessing Data Komentar.....	90
Lampiran 8. Code Fine-Tuning IndoBERT	94
Lampiran 9. Split Data Train 80%: Val 10% : Test 10%	94
Lampiran 10. Keseimbangan Data	95
Lampiran 11. Random Oversampling.....	95
Lampiran 12. Tokenisasi.....	95
Lampiran 13. Fungsi Matriks Evaluasi.....	96
Lampiran 14. Fine-Tuning IndoBERT	97
Lampiran 15. Kurva Loss & Akurasi terhadap Epoch.....	98
Lampiran 16. Model & Hasil Klasifikasi per Komentar	99
Lampiran 17. Classification Report.....	100
Lampiran 18. Confusion Matrix	100
Lampiran 19. Analisis Kesalahan Klasifikasi.....	101
Lampiran 20. Analisis Sentimen Publik terhadap Pidato Presiden	103
Lampiran 21. Penghitungan Jumlah Data & WordCloud per Kelas Sentimen.....	104