

**ANALISIS PERFORMA BERT DALAM KLASIFIKASI  
SENTIMEN PADA ISU IJAZAH PALSU**

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh Kelulusan  
Jenjang Strata Satu pada Program Studi Teknik Informatika**

**Oleh**

Erna Saputri  
2255201111



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH BENGKULU  
2026**

**LEMBAR PERSETUJUAN**

**ANALISIS PERFORMA BERT DALAM KLASIFIKASI SENTIMEN PADA ISU IJAZAH PALSU**

Oleh  
Erna Saputri  
2255201111

Tugas Akhir ini Telah Diterima dan Disahkan  
untuk Memenuhi Persyaratan Mencapai Gelar  
**SARJANA KOMPUTER (S.Kom)**

Pada  
**PROGRAM SUTDI TEKNIK INFORMATIKA**  
**FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH BENGKULU**

Bengkulu, 4 Maret 2026  
Disetujui oleh

Ketua Program Studi,

Dosen Pembimbing,

Ardi Wijaya, S.Kom., M.Kom

Ardi Wijaya, S.Kom., M.Kom

NP.19880511 201408 1 181

NIDN. 0211058803

## LEMBAR PERSETUJUAN HASIL REVISI

### ANALISIS PEFORMA BERT DALAM KLASIFIKASI SENTIMEN PADA ISU IJAZAH PALSU

Oleh

Emas Saputri  
2255101111

Telah Melakukan Revisi Sesuai dengan Perubahan  
dan Perbaikan yang Diminta Pada Saat Sidang Tugas Akhir.

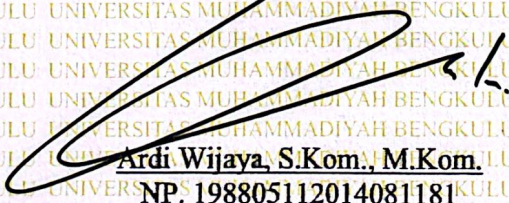
Bengkulu, 10 Maret 2026

Menyetujui

No	Nama Dosen	Keterangan	Tanda Tangan
1.	Muntahanah, S.Kom., M.Kom	Ketua Penguji	
2.	Muhammad Husni Rifqo, S.Kom., M.Kom	Penguji 1	
3.	Ardi Wijaya, S.Kom., M.Kom	Penguji 2	

Mengetahui

Ketua Program Studi Teknik Informatika

  
Ardi Wijaya, S.Kom., M.Kom.  
NP. 198805112014081181

iii

## LEMBAR PENGESAHAN

### ANALISIS PEFORMA BERT DALAM KLASIFIKASI SENTIMEN PADA ISU IJAZAH PALSU

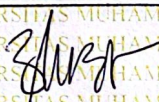
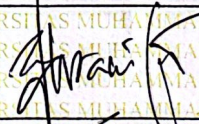

#### SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Kelulusan Jenjang Strata Satu pada Program Studi Teknik Informatika

Oleh

Erna Saputri  
2255201111

Bengkulu, 10 Maret 2026

No	Nama Dosen	Keterangan	Tanda Tangan
1.	Muntahanah, S.Kom., M.Kom	Ketua Penguji	
2.	Muhammad Husni Rifqo, S.Kom., M.Kom	Penguji 1	
3.	Ardi Wijaya, S.Kom., M.Kom	Penguji 2	

Mengesahkan

Dekan Fakultas Teknik

  
R. Guntur Alam, M.Kom., Ph.D

NP. 19730101 200004 1 040

## SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa:

- 1) Naskah Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik, baik di Universitas Muhammadiyah Bengkulu maupun perguruan tinggi lainnya.
- 2) Skripsi ini murni merupakan karya penelitian saya sendiri dan tidak menjiplak karya pihak lain. Dalam hal ada bantuan atau arahan dari pihak lain maka telah saya sebutkan identitas dan jenis bantuannya di dalam lembar ucapan terima kasih.
- 3) Seandainya ada karya pihak lain yang ternyata memiliki kemiripan dengan karya saya ini, maka hal ini adalah di luar pengetahuan saya dan terjadi tanpa kesengajaan dari pihak saya.

Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terbukti adanya kebohongan dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai norma yang berlaku di Universitas Muhammadiyah Bengkulu.

Bengkulu, 10 Maret 2026



; membuat pernyataan

Erna Saputri  
NPM. 2255201111

## MOTTO

“Jadikan kebaikan sebagai jalan hidup, karena setiap amal baik yang dilakukan dengan tulus akan mendapatkan balasan yang lebih baik dari Allah. *‘Barang siapa datang dengan kebaikan maka baginya pahala yang lebih baik dari kebaikan itu.*

’ (QS. An-Naml: 89).”

“Didikan dan perjuangan orang tuaku mengajarkanku untuk selalu berbuat baik kepada sesama, karena kebaikan adalah jalan menuju cinta Allah”

## ABSTRAK

### ANALISIS PERFORMA BERT DALAM KLASIFIKASI SENTIMEN PADA ISU IJAZAH PALSU

Nama : Erna Saputri

NPM : 2255201111

Pembimbing : Ardi Wijaya, S.Kom., M.Kom

Perkembangan teknologi informasi dan media sosial telah mengubah cara masyarakat menyampaikan dan membentuk opini publik terhadap isu sensitif, termasuk perdebatan mengenai dugaan ijazah palsu Presiden Joko Widodo. Meskipun klarifikasi resmi telah diberikan oleh institusi terkait, diskusi di ruang digital tetap memunculkan sentimen yang beragam. Analisis sentimen menjadi penting untuk memahami dinamika persepsi masyarakat secara objektif dan terukur. Metode klasifikasi konvensional seperti Naïve Bayes memiliki keterbatasan dalam memahami konteks bahasa yang kompleks, ironi, sarkasme, serta penggunaan bahasa informal di media sosial. Penelitian ini bertujuan menganalisis performa algoritma BERT dalam mengklasifikasikan sentimen masyarakat terhadap isu tersebut serta mengevaluasi akurasi, presisi, recall, dan F1-score yang dihasilkan model. Pendekatan yang digunakan adalah text mining dengan sumber data berupa cuitan dari Platform X yang dikumpulkan melalui teknik scraping menggunakan Python. Data yang diperoleh sebanyak 2.181 tweet dalam rentang waktu 2019–2025. Tahapan pengolahan data meliputi preprocessing yang terdiri atas cleansing, case folding, tokenizing, normalisasi, stopword removal, dan stemming untuk mengurangi noise serta menyeragamkan teks. Proses pelabelan dilakukan secara semi-otomatis menggunakan pendekatan lexicon-based yang divalidasi oleh anotator manusia guna meningkatkan kualitas label. Model klasifikasi yang digunakan adalah IndoBERT yang telah dilatih pada Bahasa Indonesia dan dilakukan fine-tuning untuk klasifikasi tiga kelas sentimen, yaitu positif, negatif, dan netral. Evaluasi performa dilakukan menggunakan confusion matrix serta pengukuran akurasi, presisi, recall, dan F1-score dengan pendekatan macro dan weighted average untuk mengatasi ketidakseimbangan kelas. Hasil penelitian menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan performa baik dan stabil. Nilai presisi dan recall yang seimbang menghasilkan F1-score representatif sehingga model dinilai efektif dalam menganalisis sentimen publik terhadap isu sensitif di media sosial.

**Kata Kunci : Analisis Sentimen, IndoBERT, Platform X, Deep Learning.**

## ABSTRACT

### ANALISIS PERFORMA BERT DALAM KLASIFIKASI SENTIMEN PADA ISU IJAZAH PALSU

Nama : Erna Sautri

NPM : 2255201111

Pembimbing : Ardi Wijaya, S.Kom., M.Kom

*The development of information technology and social media has transformed the way people express and shape public opinion on sensitive issues, including the debate surrounding the alleged fake diploma of President Joko Widodo. Although official clarification has been provided by the relevant institutions, discussions in the digital space continue to generate diverse sentiments. Sentiment analysis is essential to objectively and systematically understand the dynamics of public perception. Conventional classification methods such as Naïve Bayes have limitations in capturing complex linguistic contexts, irony, sarcasm, and informal language commonly found on social media. This study aims to analyze the performance of the BERT algorithm in classifying public sentiment toward the issue and to evaluate the model's accuracy, precision, recall, and F1-score. The research employs a text mining approach using tweet data collected from Platform X through a scraping technique implemented in Python. A total of 2,181 tweets were gathered within the period of 2019–2025. The data processing stages include preprocessing steps such as cleansing, case folding, tokenizing, normalization, stopword removal, and stemming to reduce noise and standardize the text. The labeling process was conducted semi-automatically using a lexicon-based approach and validated by human annotators to improve label quality. The classification model utilized in this study is IndoBERT, which has been pretrained on the Indonesian language and fine-tuned for three-class sentiment classification: positive, negative, and neutral. Model performance was evaluated using a confusion matrix along with accuracy, precision, recall, and F1-score metrics calculated using macro and weighted averages to address class imbalance. The results indicate that the model is capable of classifying sentiment with good and stable performance. Balanced precision and recall values produced a representative F1-score, demonstrating that the model is effective in analyzing public sentiment on sensitive issues in social media.*

**Keywords:** *Sentiment Analysis, IndoBERT, X Platform, Deep Learning*

## KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Performa BERT dalam Klasifikasi Sentimen pada Isu Ijazah Palsu” dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu.

Penulisan skripsi ini tidak terlepas dari berbagai tantangan dan hambatan yang dihadapi selama proses penelitian maupun penyusunan laporan. Namun berkat bantuan, dukungan, bimbingan, serta doa dari berbagai pihak, penulis akhirnya dapat menyelesaikan skripsi ini dengan sebaik-baiknya. Proses penyusunan skripsi ini juga memberikan banyak pengalaman dan pengetahuan baru bagi penulis, khususnya dalam bidang analisis sentimen dan penerapan metode BERT (Bidirectional Encoder Representations from Transformers) dalam pengolahan data teks.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penelitian dan penulisan laporan tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Susianto, M.Si selaku rektor Universitas Muhammadiyah Bengkulu
2. Bapak RG Guntur Alam, M.Kom Ph D, Selaku Dekan Fakultas Teknik Universitas Muhammadiyah Bengkulu
3. Bapak Ardi Wijaya, S.Kom., M.Kom selaku Ketua Program Studi Teknik Informatik Universitas Muhammadiyah Bengkulu
4. Bapak Ardi Wijaya, S.Kom., M.Kom selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan yang sangat berarti selama proses penelitian dan penyusunan skripsi ini hingga dapat diselesaikan dengan baik.

5. Kedua orang tua yang senantiasa memberikan doa, dukungan, semangat, serta motivasi kepada penulis selama menempuh pendidikan hingga terselesaikannya skripsi ini.
6. Rekan – rekan Mahasiswa Program Studi Teknik Informatika yang telah memberikan dukungan, bantuan, serta berbagi pengalaman selama proses perkuliahan maupun dalam penyusunan skripsi ini.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah memberikan bantuan, dukungan, dan kontribusi baik secara langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan belum sepenuhnya sempurna, baik dari segi penyusunan maupun analisis penelitian. Hal ini disebabkan oleh keterbatasan pengetahuan, pengalaman, serta waktu yang dimiliki penulis selama proses penyusunan skripsi ini. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak demi perbaikan di masa mendatang.

Penulis berharap skripsi ini dapat memberikan manfaat serta menambah wawasan bagi pembaca, khususnya dalam bidang teknologi informasi. Selain itu, skripsi ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis sentimen, pengolahan bahasa alami (Natural Language Processing), serta pemanfaatan model BERT dalam klasifikasi teks.

Wa'alaikumsalam Wr. Wb.

Bengkulu, 10 Maret 2026

Erna Saputri

## UCAPAN TERIMA KASIH

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala limpahan dan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul “Analisis Performa BERT Dalam Klasifikasi Sentimen Pada Isu Ijazah Palsu” dalam bentuk maupun isinya yang sangat sederhana.

Penulisan skripsi ini, penulis merasa masih banyak kekurangan-kekurangan baik pada teknis penulisan maupun materi. Kritik dan saran dari semua pihak sangat penulis harapkan demi penyempurnaan skripsi ini. Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, dan nasihat dari berbagai pihak selama penyusunan skripsi. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Penulis mengucapkan rasa syukur yang sebesar-besarnya kepada Allah SWT, karena atas rahmat, karunia, dan ridho-Nya penulis diberikan kekuatan, kesehatan, serta kemudahan dalam setiap langkah hingga mampu menyelesaikan skripsi ini. Dalam setiap proses yang dilalui—baik saat lelah, ragu, maupun ketika harus berjuang lebih keras—selalu ada pertolongan dan jalan yang Allah hadirkan. Semoga segala usaha yang telah dilakukan menjadi bagian dari ikhtiar yang bernilai dan senantiasa berada dalam keberkahan-Nya.
2. Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Ardi Wijaya, S.Kom., M.Kom. selaku dosen pembimbing yang telah dengan sabar meluangkan waktu, tenaga, dan pikiran dalam membimbing penulis selama proses penyusunan skripsi ini. Terima kasih atas setiap arahan, nasihat, serta ilmu yang telah diberikan, sehingga penulis dapat menyelesaikan penelitian ini dengan baik. Semoga segala kebaikan dan ilmu yang telah Bapak berikan menjadi amal yang bernilai serta membawa keberkahan.
3. Saya persembahkan karya sederhana ini kepada Bapak, sosok lelaki kuat yang selalu berjuang dalam diam. Seorang ayah yang mungkin tidak banyak mengucapkan kata, tetapi cintanya selalu terasa dalam setiap usaha dan pengorbanannya. Terima kasih, Pak, atas setiap keringat yang engkau jatuhkan, atas setiap doa yang engkau

panjatkan tanpa pernah aku dengar, dan atas setiap langkah yang engkau tempuh demi masa depan anakmu. Dari Bapak aku belajar tentang keteguhan, kesabaran, dan arti tanggung jawab dalam menjalani hidup. Karya ini mungkin tidak sebanding dengan segala perjuangan Bapak, namun melalui tulisan ini aku ingin mengucapkan terima kasih. Semoga Allah selalu menjaga langkah Bapak, memberikan kesehatan, kebahagiaan, dan umur yang panjang.

4. Tidal lupa saya juga persembahkan karya sederhana ini kepada Mama, sosok perempuan hebat yang selalu hadir dengan cinta dan doa yang tidak pernah putus. Jika Bapak adalah kekuatan yang mengajarkanku tentang keteguhan, maka Mama adalah kelembutan yang mengajarkanku tentang kasih sayang dan keikhlasan. Terima kasih, Ma, atas setiap pelukan yang menenangkan, setiap nasihat yang menguatkan, dan setiap doa yang tak pernah lelah Mama panjatkan untuk langkah-langkahku. Dari Mama aku belajar tentang kesabaran, ketulusan, dan arti cinta yang begitu luas tanpa batas. Karya sederhana ini mungkin belum mampu membalas semua pengorbanan Mama, namun melalui tulisan ini aku ingin mengucapkan terima kasih yang tulus. Semoga Allah selalu menjaga Mama, memberikan kesehatan, kebahagiaan, serta umur yang panjang.
5. Saya juga mengucapkan terima kasih kepada keluarga besar yang selalu memberikan doa, dukungan, serta semangat dalam setiap langkah perjalanan saya. Kehadiran kalian menjadi penguat di saat lelah dan menjadi pengingat bahwa saya tidak pernah benar-benar berjalan sendiri. Terima kasih atas setiap doa yang dipanjatkan, perhatian yang diberikan, dan kepercayaan yang selalu kalian titipkan kepada saya. Semoga kebersamaan dan kasih sayang dalam keluarga ini selalu terjaga, serta menjadi sumber kekuatan dalam setiap perjalanan hidup yang akan datang.
6. Saya juga mengucapkan terima kasih kepada seseorang yang telah kebersamai perjalanan ini, Seseorang yang memiliki nama Gading Cempaka yang selama kurang lebih tiga tahun terakhir berjalan bersama penulis, melewati berbagai cerita yang penuh warna, baik suka maupun duka. Terima kasih karena telah tetap hadir di tengah proses yang tidak selalu mudah. Dalam perjalanan waktu itu, ada tawa yang dibagi, lelah yang saling dikuatkan, serta doa dan semangat yang terus saling diberikan.

Kehadiranmu menjadi salah satu bagian dari perjalanan ini, menjadi pengingat bahwa setiap langkah terasa lebih ringan ketika dijalani bersama. Semoga waktu yang telah dilalui menjadi cerita baik yang selalu dikenang, dan semoga setiap langkah ke depan selalu dipenuhi dengan kebaikan dan harapan yang indah.

7. Saya juga mengucapkan terima kasih kepada seorang perempuan yang telah saya kenal sejak masa sekolah menengah pertama, yaitu Bening Nawang Arum Hapsari, yang hingga hari ini tetap berjalan bersama dalam perjalanan hidup saya. Terima kasih karena telah menjadi bagian dari hari-hari sederhana yang ternyata begitu berarti—menjadi teman dalam tawa, dalam cerita panjang di malam hari, dan dalam langkah-langkah kecil kehidupan yang kita jalani bersama. Dari hal-hal yang tampak biasa itulah tersimpan kenangan yang begitu berharga. Terima kasih karena telah menjadi saksi perjalanan penulis, menyaksikan setiap tangis yang jatuh diam-diam, setiap doa yang dipanjatkan dengan penuh harap, serta setiap usaha yang terus diperjuangkan hingga sampai pada titik ini. Kehadiranmu adalah bagian dari cerita yang tidak tergantikan. Semoga persahabatan ini tetap tumbuh bersama waktu, tetap hangat dalam kebersamaan, dan selalu menjadi rumah kecil bagi setiap cerita yang kita simpan.
8. Saya juga mengucapkan terima kasih kepada Bedha Diyah Utami Fatmawati dan Ratna Juwita, sosok yang saya kenal sejak pertama kali menginjak bangku perkuliahan. Terima kasih karena selama hampir empat tahun ini telah menjadi teman yang baik dan turut membersamai berbagai proses dalam perjalanan hidup penulis. Dalam waktu yang tidak sebentar itu, banyak cerita yang telah dilalui bersama—berbagi tawa, cerita, dan saling menguatkan di tengah perjalanan yang tidak selalu mudah. Kehadiranmu menjadi salah satu bagian dari kenangan selama masa perkuliahan yang begitu berarti. Terima kasih atas setiap kebaikan, dukungan, serta kebersamaan yang telah diberikan. Semoga persahabatan ini tetap terjaga dengan baik, dan semoga setiap langkah ke depan selalu dipenuhi dengan kebaikan, kebahagiaan, serta keberhasilan dalam perjalanan hidupmu.
9. Penulis juga mengucapkan terima kasih kepada teman-teman lainnya yang tidak dapat disebutkan satu per satu. Terima kasih karena telah menjadi bagian dari

perjalanan ini, mengisi hari-hari dengan kebersamaan, cerita, serta dukungan yang begitu berarti. Setiap tawa, percakapan, dan kenangan yang tercipta selama perjalanan ini menjadi bagian dari kisah yang tidak akan terlupakan. Semoga kebersamaan yang pernah kita lalui menjadi cerita indah yang selalu dikenang, dan semoga langkah kita ke depan selalu dipenuhi dengan kebaikan, keberhasilan, dan kebahagiaan.

10. Terakhir, izinkan saya mengucapkan terima kasih kepada satu sosok perempuan yang selama ini diam-diam berjuang tanpa banyak suara. Seorang gadis sederhana yang mungkin terlihat biasa saja, namun menyimpan kekuatan yang tidak semua orang mampu pahami. Terima kasih kepada peneliti skripsi ini, yaitu diriku sendiri, Erna Saputri, anak perempuan terakhir sekaligus salah satu harapan bagi kedua orang tuanya. Terima kasih karena telah hadir di dunia ini, karena telah bertahan sejauh ini, dan karena tidak pernah benar-benar menyerah meskipun langkahmu sering terasa berat. Terima kasih karena tetap berjalan, meskipun pelan, meskipun lelah, dan meskipun terkadang harus melalui jalan yang sepi tanpa banyak tepuk tangan dari orang lain. Aku bangga pada setiap langkah kecil yang kau ambil. Pada setiap air mata yang kau sembunyikan, pada setiap kegagalan yang kau ubah menjadi pelajaran, dan pada setiap keberanian untuk tetap melanjutkan perjalanan. Tidak semua perjuangan terlihat oleh dunia, tetapi setiap proses yang kau jalani adalah bukti bahwa kamu lebih kuat dari yang kamu kira. Ketika harapan tidak selalu sejalan dengan apa yang semesta berikan, belajarlah untuk tetap bersyukur. Belajarlah menerima bahwa hidup tidak selalu tentang mendapatkan apa yang diinginkan, tetapi tentang memahami apa yang telah diberikan. Tetaplah menjadi pribadi yang baik, bahkan ketika dunia terkadang tidak berlaku demikian. Tetaplah memiliki hati yang lembut untuk memaafkan, hati yang teduh dalam perkataan dan perbuatan, serta jiwa yang selalu berusaha menebarkan kebaikan kepada siapa pun yang kau temui. Jangan pernah lelah untuk berusaha dan jangan lupa untuk tetap berbahagia. Rayakan setiap bagian dari dirimu, setiap langkah kecil yang membawa dirimu semakin dekat pada mimpi-mimpimu. Jadikan dirimu seseorang yang bermanfaat, tidak hanya bagi dirimu sendiri, tetapi juga bagi orang lain di sekitarmu. Aku berdoa semoga setiap

langkah kecilmu selalu diperkuat, dipertemukan dengan orang-orang baik yang menjaga dan mendukung perjalananmu, serta setiap mimpi yang kau simpan perlahan menemukan jalannya untuk menjadi nyata. Semoga hatimu selalu dijaga agar tetap tulus, pikiranmu tetap kuat, dan langkahmu tetap teguh dalam menapaki perjalanan hidup yang masih panjang.

Untuk diriku sendiri, terima kasih karena tidak pernah benar-benar berhenti mencoba. Terima kasih karena tetap memilih untuk bangkit, kembali berjalan, dan terus percaya bahwa suatu hari nanti semua perjuangan ini akan menemukan maknanya.

## DAFTAR ISI

COVER .....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PERSETUJUAN HASIL REVISI.....	iii
LEMBAR PENGESAHAN.....	iv
SURAT PERNYATAAN.....	v
DAFTAR RIWAYAT HIDUP.....	vi
MOTTO.....	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
KATA PENGANTAR.....	x
UCAPAN TERIMA KASIH.....	xii
DAFTAR ISI.....	xvii
DAFTAR TABEL.....	xix
DAFTAR GAMBAR.....	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Kerangka Penelitian.....	5
BAB II LANDASAN TEORI.....	2
2.1 Tinjauan Pustaka.....	2
2.2 Data Mining.....	12
2.3 Pengelompokan Data.....	13
2.4 Proses Data Mining.....	14
2.5 Machine Learning.....	15

2.6 Deep Learning .....	18
2.7 Text Mining .....	19
2.8 Text Preprocessing .....	20
2.9 Algoritma BERT .....	23
2.10 Platform X.....	31
<b>BAB III ANALISIS MASALAH DAN PERANCANGAN PROGRAM .....</b>	<b>33</b>
3.1 Framework .....	33
3.2 Metode Pengumpulan Data.....	39
3.3 Bahan Penelitian.....	42
<b>BAB IV IMPLEMENTASI DAN UJI COBA .....</b>	<b>44</b>
4.1 Pengambilan Data .....	44
4.2 Text Preprocessing .....	45
4.3 Pelabelan Leksikon.....	55
4.4 Ekstraksi Opini.....	58
4.5 Evaluasi Model BERT .....	60
4.6 Evaluasi Manual Model BERT .....	72
<b>BAB V PENUTUP .....</b>	<b>71</b>
5.1 Kesimpulan .....	71
5.2 Saran.....	72
<b>DAFTAR PUSTAKA.....</b>	<b>74</b>

## DAFTAR TABEL

Tabel 2. 1 Literatur Review .....	9
Tabel 3. 1 Framework.....	33
Tabel 3. 2 Bahan Penelitian.....	42
Tabel 4. 1 Hasil Cleansing Data .....	46
Tabel 4. 2 Hasil <i>Case Folding</i> .....	47
Tabel 4. 3 Hasil Tokenizing .....	48
Tabel 4. 4 Tabel Kamus Alay.....	50
Tabel 4. 5 Hasil Normalisasi Data.....	52
Tabel 4. 6 Hasil <i>Stopword</i> .....	53
Tabel 4. 7 Hasil <i>Stemming</i> .....	54
Tabel 4. 8 Contoh Hasil Label <i>Lexicon</i> .....	56

## DAFTAR GAMBAR

Gambar 1. 1 Kerangka Penelitian.....	6
Gambar 2. 1 Pre-training dan Fine-tuning pada BERT.....	26
Gambar 2. 2 Konstruksi Input/Output BERT.....	28
Gambar 4. 1 Hasil Scraping Data.....	45
Gambar 4. 2 Hasil Ekstraksi Opini.....	58
Gambar 4. 3 Hasil <i>Word Cloud</i> .....	60
Gambar 4. 4 Grafik <i>Train Dataset</i> .....	62
Gambar 4. 5 Uji Coba Model.....	62
Gambar 4. 6 Confusion Matrix Result Data Testing.....	63
Gambar 4. 7 Data Prediksi Negatif dan true Label Negatif (1).....	65
Gambar 4. 8 Data Prediksi Negatif dan true Label Negatif (2).....	65
Gambar 4. 9 Data Prediksi Negatif dan true Label Negatif (3).....	66
Gambar 4. 10 Data Prediksi Negatif dan true Label Negatif (4).....	66
Gambar 4. 11 Data Prediksi Negatif dan true Label Negatif (5).....	67
Gambar 4. 12 Data Prediksi Negatif dan true Label Negatif (6).....	67
Gambar 4. 13 N Data Prediksi Negatif dan true Label Negatif (7).....	68
Gambar 4. 14 Data Prediksi Negatif dan true Label Negatif (8).....	68
Gambar 4. 15 Data Prediksi Negatif dan true Label Positif (1).....	68
Gambar 4. 16 Data Prediksi Negatif dan true Label Positif (2).....	69
Gambar 4. 17 Data Prediksi Negatif dan true Label Neutral.....	69
Gambar 4. 18 Data Prediksi Positif dan true Label Negatif.....	70

Gambar 4. 19 Data Prediksi Positif dan true Label Positif.....	70
Gambar 4. 20 Data Prediksi Positif dan true Label Neutral .....	71
Gambar 4. 21 Data Prediksi Neutral dan true Label Negatif.....	71
Gambar 4. 22 Data Prediksi Neutral dan true Label Positif .....	71
Gambar 4. 23 Data Prediksi Neutral dan true Label Neutral.....	71
Gambar 4. 24 Nilai Akurasi, Recall, Presisi dan F1-Score Google Collab .....	72

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan merupakan pilar pembangunan negara, sehingga kredibilitas lembaga pendidikan dan keaslian dokumen akademik seperti ijazah menjadi isu penting bagi kepercayaan publik. Di Indonesia, perdebatan mengenai keabsahan ijazah Presiden Joko Widodo telah berlangsung lama dan terus bergulir meskipun Universitas Gadjah Mada telah menyatakan bahwa ijazah dan skripsi beliau adalah asli (Press, U. G. M., 2025). Isu ini mencerminkan bagaimana masyarakat merespon informasi sensitif terkait legitimasi akademik dan bagaimana narasi dapat memengaruhi persepsi publik dalam konteks pendidikan, sosial, hingga politik.

Perkembangan teknologi informasi dan media sosial telah mempercepat penyebaran opini, hoaks, dan narasi publik, sehingga analisis sentimen menjadi penting untuk memahami dinamika persepsi masyarakat (Afifa et al., 2023). Metode NLP modern memungkinkan analisis teks dalam jumlah besar secara lebih dalam (Agustina et al., 2022). BERT, sebagai model NLP mutakhir, menawarkan kemampuan memahami konteks dua arah dan menangkap nuansa bahasa seperti ironi atau sarkasme, sehingga lebih efektif digunakan pada isu yang sarat opini dan ambiguitas (Chandradev et al., 2023; Doloksaribu & Samuel, 2022; Kusnadi et al., 2021).

Analisis sentimen sendiri telah banyak dilakukan, misalnya pada penelitian *Analisis Sentimen Opini Masyarakat Terhadap Presiden Jokowi Sebelum dan Sesudah*

*Pilpres 2024 Menggunakan Metode Naive Bayes Classification*, yang menunjukkan bagaimana metode Naive Bayes mampu mengelompokkan opini publik menjadi sentimen positif, negatif, dan netral berdasarkan data media sosial. Penelitian tersebut juga menunjukkan peningkatan akurasi hingga 64%, dengan precision tertinggi pada sentimen netral (88%) dan recall tertinggi pada sentimen positif (87%), namun metode tersebut masih menghadapi keterbatasan dalam memahami konteks bahasa yang ambigu, sarkastik, maupun menggunakan bahasa gaul. W2

Namun demikian, metode klasifikasi sentimen klasik seperti Naive Bayes memiliki sejumlah keterbatasan mendasar, terutama dalam memahami konteks bahasa yang kompleks. Naive Bayes bekerja dengan asumsi independensi antar kata, sehingga makna sebuah kalimat dipandang sebagai akumulasi kata-kata individual tanpa mempertimbangkan hubungan semantik dan konteks kalimat secara utuh. Pendekatan ini membuat metode tersebut kurang efektif dalam menangani ironi, sarkasme, negasi ganda, serta penggunaan bahasa gaul atau slang yang lazim digunakan di media sosial. Pada isu sensitif seperti tuduhan “ijazah palsu”, banyak opini masyarakat disampaikan secara implisit, menyindir, atau menggunakan humor sarkastik, yang sering kali diklasifikasikan secara keliru oleh metode berbasis probabilistik sederhana.

Selain itu, evaluasi kinerja metode klasik sering menunjukkan ketidakseimbangan antara metrik akurasi, presisi, dan recall. Akurasi yang terlihat cukup tinggi tidak selalu mencerminkan kemampuan model dalam mengklasifikasikan setiap kelas sentimen secara adil, terutama ketika distribusi data tidak seimbang. Naive Bayes cenderung menghasilkan presisi atau recall yang timpang pada kelas tertentu, sehingga nilai F1-score sebagai metrik keseimbangan sering kali rendah atau tidak stabil. Hal ini berdampak pada

kurang dapat diandalkan hasilnya dalam menggambarkan persepsi publik secara komprehensif, khususnya ketika opini bersifat ambigu dan berlapis.

Oleh karena itu, diperlukan pendekatan yang mampu melakukan pemodelan konteks bahasa secara lebih mendalam dan dievaluasi menggunakan metrik kinerja yang lebih representatif, seperti akurasi, presisi, recall, dan F1-score secara simultan. Penggunaan metrik-metrik tersebut memungkinkan penilaian yang lebih objektif terhadap kemampuan model dalam memahami sentimen publik secara menyeluruh, bukan hanya dari sisi ketepatan prediksi, tetapi juga konsistensi dan keseimbangan klasifikasi antar kelas sentimen. Dalam konteks ini, BERT menjadi solusi yang relevan karena mampu mengatasi keterbatasan metode klasik dan menghasilkan analisis sentimen yang lebih andal dalam memahami persepsi publik terhadap isu pendidikan yang sensitif dan kontroversial.

Untuk mengatasi hal ini, BERT menawarkan pemrosesan konteks berbasis transformer yang lebih akurat dalam menangkap ironi, sarkasme, dan bahasa informal yang sering muncul di media sosial (Aftan & Shah, 2023). Berbagai studi menunjukkan bahwa BERT unggul dalam akurasi, presisi, recall, dan F1-score dibandingkan metode konvensional seperti Naïve Bayes, terutama pada dataset dengan keragaman linguistik tinggi dan struktur kalimat yang kompleks (Irmawan et al., 2024). Keunggulan ini membuat BERT lebih potensial dalam menghasilkan analisis sentimen yang tidak hanya akurat tetapi juga lebih stabil pada isu sensitif, termasuk perdebatan mengenai “ijazah palsu Jokowi,” yang kerap memunculkan opini berlapis dan ambigu.

Penggunaan BERT relevan karena kemampuannya memahami konteks bahasa percakapan digital secara dua arah dan menangkap relasi semantik antar kata secara

mendalam (Wang et al., 2020). Analisis ini bertujuan untuk menilai efektivitas BERT dalam mengatasi keterbatasan metode klasik, khususnya dalam menangani sarkasme, ironi, ambiguitas, dan bahasa informal, sekaligus meningkatkan kinerja model berdasarkan akurasi, presisi, recall, dan konsistensi klasifikasi, sehingga hasil analisis menjadi lebih dapat diandalkan.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka perumusan masalah dalam penelitian ini adalah:

1. Bagaimana model *BERT* dapat digunakan untuk menganalisis sentimen masyarakat terhadap isu ijazah palsu Jokowi pada media sosial dan platform daring?
2. Sejauh mana tingkat akurasi model *BERT* dalam membedakan sentimen positif, negatif, dan netral terhadap isu tersebut?
3. Bagaimana pengaruh ketimpangan jumlah data antar kelas (*class imbalance*) terhadap hasil klasifikasi sentimen yang dilakukan menggunakan algoritma *BERT*?

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan algoritma *BERT* untuk melakukan analisis sentimen masyarakat terhadap isu ijazah palsu Jokowi berdasarkan data teks dari media sosial atau sumber daring.