

**ANALISIS SENTIMEN KOMENTAR INSTAGRAM TERHADAP
CALON PRESIDEN (CAPRES) 2024 MENGGUNAKAN MODEL
INDOBERT DAN INDOBERTWEET**

TUGAS AKHIR

Diajukan Oleh :

FIRA ELJA SABIDRA

NIM. 210705021

**Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi**



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI AR-RANIRY
BANDA ACEH
2025 M / 1446 H**

LEMBAR PERSETUJUAN

ANALISIS SENTIMEN KOMENTAR INSTAGRAM TERHADAP CALON PRESIDEN (CAPRES) 2024 MENGGUNAKAN MODEL INDOBERT DAN INDOBERTWEET

TUGAS AKHIR

Diajukan Kepada Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Ar-Raniry Banda Aceh
Sebagai Salah Satu Beban Studi Memperoleh Gelar Sarjana
pada Prodi Teknologi Informasi

Oleh:

Fira Elja Sabidra

NIM. 210705021

**Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi**

Disetujui Untuk Dimunaqasyahkan Oleh:

Pembimbing I,



Dr. Hendri Ahmadian, M.I.M

NIP.198301042014031002

Pembimbing II,



Raihan Islamadina, M.T.

NIP.198901312020122011

Mengetahui,

Ketua Program Studi Teknologi Informasi



Malahayati, M.T

NIP.198301272015032003

LEMBAR PENGESAHAN

ANALISIS SENTIMEN KOMENTAR INSTAGRAM TERHADAP CALON PRESIDEN (CAPRES) 2024 MENGGUNAKAN MODEL INDOBERT DAN INDOBERTWEET

TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S-1)
Dalam Program Studi Teknologi Informasi

Pada Hari/Tanggal: Selasa, 17 Mei 2025
21 Dzulhijjah 1446 H
Di Darussalam, Banda Aceh
Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,


Dr. Hendri Ahmadian, M.I.M
NIP.198301042014031002

Sekretaris,


Raihan Islamadina, M.T.
NIP.198901312020122011

Penguji I,


Khafran AR, M.Kom
NIP.198607042014031001

Penguji II,


Baihaqi, M. T
NIP.198802212022031001

Mengetahui:

Dekan Fakultas Sains dan Teknologi
UIN Ar-Raniry Banda Aceh,



Prof. Dr. Ir. Muhammad Dirhamsyah, M.T., I.P.U
NIP.196210021988111001

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Fira Elja Sabidra
NIM : 210705021
Program Studi : Teknologi Informasi
Fakultas : Sains dan Teknologi
Judul Tugas Akhir : Analisis Sentimen Komentar Instagram Terhadap Calon Presiden (Capres) 2024 Menggunakan Model Indobert dan Indobertweet

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggung jawabkan.
2. Tidak melakukan plagiasi terhadap naskah orang lain.
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya.
4. Tidak memanipulasi dan memalsukan data.
5. Mengerjakan sendiri karya ini dan mampu bertanggung jawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggung jawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenakan sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh.

Demikian Pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

Banda Aceh, 17 Juni 2025

Yang Menyatakan



2000
METERAI
TEMPEL
53AMX327235022

Fira Elja Sabidra

ABSTRAK

Nama : Fira Elja Sabidra
NIM : 210705021
Program Studi : Teknologi Informasi
Judul : Analisis Sentimen Komentar Instagram Terhadap Calon Presiden (Capres) 2024 Menggunakan Model IndoBERT dan IndoBERTweet.
Tanggal Sidang : 17 Juni 2025
Jumlah Halaman : 88
Pembimbing I : Dr. Hendri Ahmadian, M.I.M
Pembimbing II : Raihan Islamadina, M.T
Kata Kunci : Analisis sentimen, Instagram, Capres 2024, IndoBERT, IndoBERTweet.

Instagram menjadi salah satu platform utama bagi masyarakat untuk menyampaikan opini terkait calon presiden (capres) menjelang Pemilu 2024. Penelitian ini bertujuan untuk menganalisis sentimen komentar Instagram terhadap capres 2024 dengan memanfaatkan empat model berbasis Transformer: IndoBERT v1, IndoBERT v2, IndoBERTweet v1, dan IndoBERTweet v2. Data dikumpulkan melalui proses *web scraping*, kemudian dilakukan labeling, pra-pemrosesan, dan tokenisasi. Evaluasi dilakukan menggunakan metrik *akurasi*, *precision*, *recall*, dan *f1-score* (*macro* dan *weighted*). Hasil menunjukkan bahwa IndoBERT v2 merupakan model dengan performa terbaik, memperoleh *akurasi* sebesar 0,84 dan *f1-score macro* sebesar 0,82. Model IndoBERTweet v2 berada di posisi kedua dengan *akurasi* 0,81 dan *f1-score macro* 0,80, menunjukkan efektivitasnya dalam memahami bahasa informal di media sosial. Sebaliknya, IndoBERT v1 dan IndoBERTweet v1 menunjukkan performa lebih rendah. Hasil ini menegaskan bahwa pemilihan model terbaik perlu mempertimbangkan tidak hanya metrik evaluasi, tetapi juga kesesuaian model terhadap karakteristik data yang digunakan.

Kata Kunci: Analisis sentimen, Instagram, Capres 2024, IndoBERT, IndoBERTweet.

ABSTRACT

Name : Fira Elja Sabidra
NIM : 210705021
Study Program : Information Technology
Title : Analysis of Instagram Comment Sentiment Towards the 2024 Presidential Candidates Using the IndoBERT and IndoBERTweet Models.
Trial Date : June 17, 2025
Number of pages : 88
Supervisor I : Dr. Hendri Ahmadian, M.I.M
Supervisor II : Raihan Islamadina, M.T
Keywords : Sentiment analysis, Instagram, 2024 Presidential Candidate, IndoBERT, IndoBERTweet.

Instagram has become one of the main platforms for people to express their opinions regarding presidential candidates ahead of the 2024 Election. This study aims to analyze the sentiment of Instagram comments on the 2024 presidential candidates by utilizing four Transformer-based models: IndoBERT v1, IndoBERT v2, IndoBERTweet v1, and IndoBERTweet v2. Data was collected through a web scraping process, then labeling, pre-processing, and tokenization were carried out. Evaluation was carried out using accuracy, precision, recall, and f1-score (macro and weighted) metrics. The results show that IndoBERT v2 is the model with the best performance, obtaining an accuracy of 0.84 and a macro f1-score of 0.82. The IndoBERTweet v2 model is in second place with an accuracy of 0.81 and a macro f1-score of 0.80, indicating its effectiveness in understanding informal language on social media. In contrast, IndoBERT v1 and IndoBERTweet v1 show lower performance. These results confirm that the selection of the best model needs to consider not only the evaluation metrics, but also the suitability of the model to the characteristics of the data used.

Keywords: Sentiment analysis, Instagram, 2024 Presidential Candidate, IndoBERT, IndoBERTweet.

KATA PENGANTAR

Bismillahirrahmanirrahim, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir yang berjudul “Aanalisis Sentimen Komentar Instagram Terhadap Calon Presiden (Capres) 2024 Menggunakan Model IndoBERT dan IndoBERTweet”. Shalawat dan salam semoga selalu tercurah kepada junjungan Nabi Muhammad SAW, keluarga, para sahabat, dan seluruh pengikut beliau hingga akhir zaman, yang telah membawa umat manusia dari zaman kegelapan menuju cahaya ilmu dan kebenaran.

Dalam menyelesaikan penyusunan tugas akhir ini, penulis menyadari bahwa proses ini tidak akan dapat terlaksana dengan baik tanpa adanya dukungan, arahan, semangat, serta bantuan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua tercinta sekaligus pintu menuju surga, Ibunda Eli Yeni Rahmi dan Ayah Hindra Mailisa S.K.M terima kasih atas kasih sayang yang tiada batas, doa yang selalu menyertai, serta dukungan luar biasa yang tak pernah surut. Tanpa kalian, langkah putrimu takkan sampai sejauh ini.
2. Untuk saudara tercinta, Abang, Kakak dan Adik perempuanku tersayang, serta kesayangan embik Aca Alik terima kasih atas doa, semangat, dan kehadiran kalian yang menjadi tempat bercerita sekaligus penyemangat. Segala dukungan kalian telah menjadi bagian penting dalam terselesaikannya tugas akhir ini.
3. Bapak Dr. Hendri Ahmadian, M.I.M. selaku Pembimbing I dan Ibu Raihan Islamadina, M.T. selaku Pembimbing II, penulis mengucapkan terima kasih yang sebesar-besarnya atas bimbingan, masukan, serta waktu yang telah diluangkan di tengah kesibukan. Menjadi bagian dari bimbingan Bapak dan Ibu adalah sebuah anugerah yang tak ternilai. Semoga segala kebaikan dan dedikasi dibalas dengan limpahan kesehatan dan keberkahan.

4. Bapak Prof. Dr. Ir. M. Dirhamsyah, M.T., IPU selaku Dekan Fakultas Sains dan Teknologi UIN Ar-Raniry.
5. Ibu Malayahati, M.T. dan Bapak Khairan AR, M.Kom. selaku Ketua dan Sekretaris Prodi Teknologi Informasi, yang telah memberikan bimbingan dan arahan dalam menyelesaikan tugas akhir ini.
6. Bapak Dr. Hendri Ahmadian, M.I.M dan bapak Mulkan Fadhli, S.T., M.T.sebagai Penasehat Akademik yang senantiasa memberikan arahan, motivasi dan nasehat kepada penulis dari semester 1 hingga semester 8. Semoga semua yang bapak berikan kepada saya dibalaskan di akhirat kelak.
7. Ibu Cut Ida Rahmadiana, S.Si. selaku Staff Prodi Teknologi Informasi, yang senantiasa membantu dan selalu sabar menjelaskan kepada penulis dalam pemberkasan administrasi.
8. Bapak dan Ibu dosen Program Studi Teknologi Informasi yang telah membekali penulis dengan ilmu pengetahuan di bidang Teknologi Informasi.
9. Kepada sahabat-sahabat surga yaitu Fadila Khairani, Nuri Harpiyani, Yenni Nurul Hasanah, Ulfa Fitri, Azlina Permaynuri, Era Syafina, Cut Raudhatul Ilmi dan Dina Andriana terima kasih telah menjadi tempat berbagi cerita, pelipur lara, dan rumah pertama penulis di tanah rantau ini. Kita telah melalui perjuangan ini bersama, dan kenangan itu akan selalu penulis kenang sepanjang perjalanan hidup. Semoga kita semua selalu dalam lindungan Allah SWT.
10. Kepada teman-teman seperjuangan yang selalu mendukung dan memberi semangat yaitu Filahi Akbar, Hanafi Akbar dan Muhammad Uzir. Semoga segala hal-hal baik menyertai kalian juga.
11. Kepada teman-teman Letting 21 Teknologi Informasi, yang telah menemani penulis dari semester 1 hingga 8. Terima kasih atas kebersamaan, dukungan, dan semangat yang selalu diberikan. Semoga ilmu yang kita peroleh bermanfaat dan mengantarkan kita menuju kesuksesan di masa depan.
12. Pihak-pihak terkait yang lainnya yang membantu penulis dalam penyusunan tugas akhir ini.

13. Terakhir, terima kasih kepada wanita sederhana yang memiliki impian besar, meskipun terkadang sulit dimengerti isi kepalanya, diriku sendiri, Fira Elja Sabidra. Terima kasih telah berusaha keras untuk meyakinkan dan menguatkan diri bahwa aku mampu menyelesaikan studi ini hingga akhir. Berbahagialah selalu dengan dirimu sendiri, Fira. Rayakan kehadiranmu sebagai berkah di setiap langkah yang kamu ambil. Jangan sia-siakan usaha dan doa yang selalu kamu panjatkan. Allah telah merencanakan yang terbaik dan memberikan porsi terbaik dalam setiap perjalanan hidupmu. Semoga langkah-langkah kebaikan selalu menyertaimu, dan semoga Allah senantiasa meridhai setiap langkahmu serta melindungimu. Aamiin.

Penulis menyadari sepenuhnya bahwa dalam penyusunan tugas akhir ini tidak cukup dikategorikan sempurna, untuk itu penulis dengan segala kerendahan hati menerima saran dan kritikan dalam menyempurnakan penyusunan tugas akhir ini. Akhir kata, semoga tugas akhir ini dapat memberikan manfaat bagi penulis juga pembaca dan semoga dicatat sebagai sebuah amal kebaikan oleh Allah SWT. *Aamiin Aamiin Ya Rabbal A'lam.*

Banda Aceh, 17 Juni 2025

Penulis

Fira Elja Sabidra

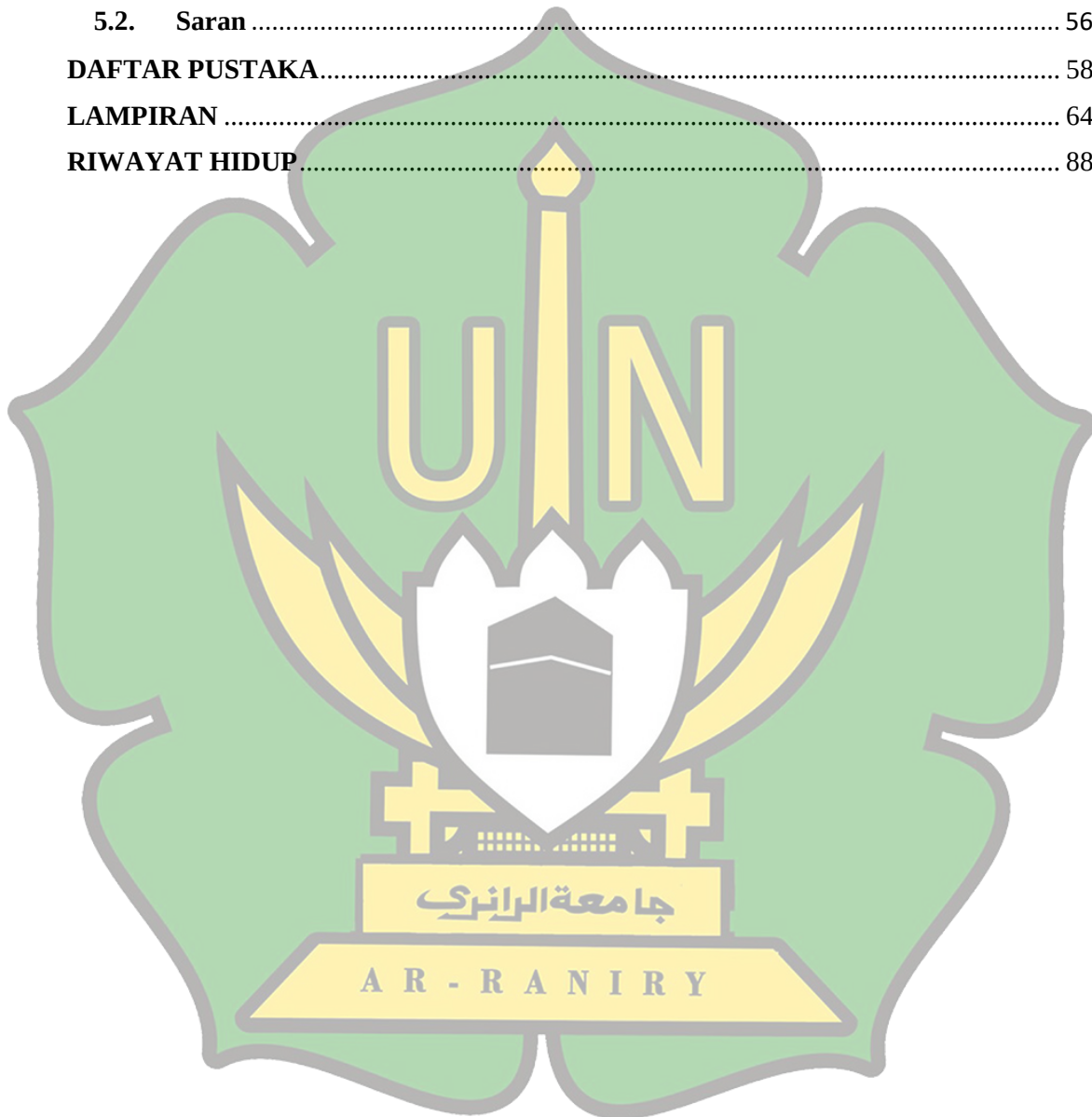


DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN KEASLIAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TAEBL	xiii
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Manfaat Penelitian	4
BAB II	5
TINJAUAN PUSTAKA	5
2.1. Penelitian Terdahulu	5
2.2. Analisis Sentimen	10
2.3. <i>Natural Language Processing</i> (NLP)	11
2.4. Bidirectional Encoder Representations from Transformers (BERT)	11
2.5. IndoBERT	13
2.6. IndoBERTweet	14
2.7. IndoLEM	15
2.8. Web Scraping	15
2.9. Instagram	16
2.10. Tokoh Politik	17
2.11. Evaluasi Model	18
2.11.1. Confucion Matrix	18

2.11.2.	Grafik Akurasi dan Loss	20
2.12.	Tools	21
2.13.	Kerangka Berfikir Penelitian.....	22
BAB III.....		23
METODOLOGI PENELITIAN		23
3.1.	Tahap penelitian.....	23
3.2.	Studi Literatur	24
3.3.	Pengumpulan Data.....	24
3.4.	Pelabelan Data.....	24
3.5.	Data Preprocessing	24
3.6.	Split Data	26
3.7.	Load Model.....	26
3.8.	<i>Fine-tuning</i>	26
3.9.	Evaluasi Model.....	27
3.10.	Perbandingan Model.....	28
3.11.	Lokasi Penelitian	28
3.12.	Alat Bantu Penelitian	28
BAB IV.....		29
HASIL DAN PEMBAHASAN.....		29
4.1.	Pengumpulan Data.....	29
4.2.	Pelabelan Data.....	30
4.3.	Hasil Preprocessing	32
4.4.	Split dataset	34
4.5.	Fine – Tuning.....	35
4.6.	Load Model.....	36
4.7.	Evaluasi Model.....	38
4.7.1.	Visualisasi Data	38
4.7.3.	Grafik Akurasi Train Loss dan Confusion Matrix	41
4.7.3.1.	IndoBER Versi Pertama (v1)	42
4.7.3.2.	IndoBERT Versi Kedua (v2).....	44
4.7.3.3.	IndoBERTweet Versi Pertama (v1).....	46
4.7.3.4.	IndoBERTweet Versi Kedua (v2)	48

4.7.4. Clasification Report	50
4.8. Perbandingan Model	52
BAB V	56
KESIMPULAN DAN SARAN	56
5.1. Kesimpulan	56
5.2. Saran	56
DAFTAR PUSTAKA.....	58
LAMPIRAN	64
RIWAYAT HIDUP.....	88

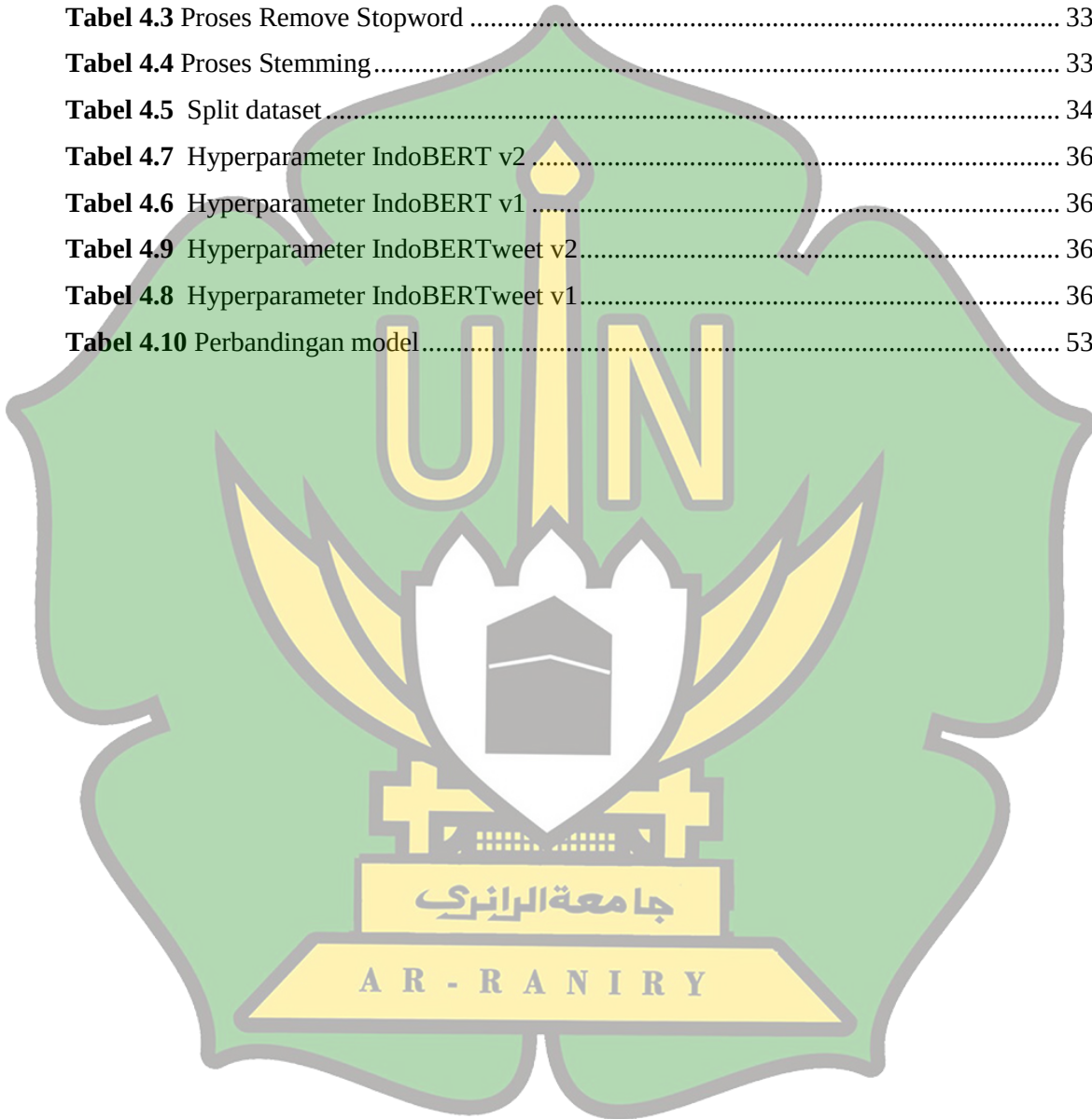


DAFTAR GAMBAR

Gambar 2.1 Arsitektur BERT (Devlin et al., 2018)	12
Gambar 2.2 Representasi Input BERT (Devlin et al., 2018).....	12
Gambar 2.3 Mekanisme Pre-training dan Fine-tuning (Devlin et al., 2018).....	13
Gambar 2.4 langkah web scraping	16
Gambar 2.5 confusion matrix.....	19
Gambar 2.6 kerangka berfikir	22
Gambar 3.1 alur tahap penelitian	23
Gambar 3.2 Tahapan pre-train dan finetuning	27
Gambar 4.1 Komentar pengguna instagram pada postingan akun politik.....	29
Gambar 4.2 Komentar dari proses web scraping.....	30
Gambar 4.3 Komentar yang sudah terlabel	31
Gambar 4.4 Pie chart sentimen.....	31
Gambar 4.5 Code Penerapan model IndoBERT.....	37
Gambar 4.6 Code Penerapan model IndoBERTweet.....	38
Gambar 4.7 WordCloud Positif.....	39
Gambar 4.8 WordCloud Negatif.....	39
Gambar 4.9 WordCloud Netral	40
Gambar 4.10 Word Distribution.....	41
Gambar 4.11 Grafik Train & Validation IndoBert (v1)	42
Gambar 4.12 Confusion matrix IndoBert (v1)	43
Gambar 4.13 Grafik Train & Validation IndoBert (v2)	44
Gambar 4.14 Confusion matrix IndoBert (v2)	45
Gambar 4.15 Grafik Train & Validation IndoBertweet (v1).....	46
Gambar 4.16 Confusion matrix IndoBertweet (v1).....	47
Gambar 4.17 Grafik Train & Validation IndoBertweet (v2).....	48
Gambar 4.18 Confusion matrix IndoBertweet (v2).....	49
Gambar 4.19 Performa model IndoBERT v1 dengan Batch size 32 dan epoch 20	50
Gambar 4.20 Performa model IndoBERT v2 dengan Batch size 16 dan epoch 10	51
Gambar 4.21 Performa model IndoBERTweet v1 dengan Batch size 32 dan epoch 20	51
Gambar 4.22 Performa model IndoBERTweet v2 dengan Batch size 16 dan epoch 10	52

DAFTAR TAEBL

Tabel 2.1 Perbandingan Penelitian Sebelumnya	7
Tabel 4.1 Proses Data cleaning dan Case folding	32
Tabel 4.2 Proses Normalisasi	32
Tabel 4.3 Proses Remove Stopword	33
Tabel 4.4 Proses Stemming.....	33
Tabel 4.5 Split dataset.....	34
Tabel 4.7 Hyperparameter IndoBERT v2	36
Tabel 4.6 Hyperparameter IndoBERT v1	36
Tabel 4.9 Hyperparameter IndoBERTweet v2.....	36
Tabel 4.8 Hyperparameter IndoBERTweet v1.....	36
Tabel 4.10 Perbandingan model.....	53



BAB I

PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen adalah proses otomatis untuk memahami dan mengidentifikasi sentimen dalam opini, apakah *positif*, *negatif*, atau *netral*. Juga dikenal sebagai *opinion mining*, analisis ini menggunakan NLP (*Natural Language Processing*) untuk mengevaluasi pendapat dan emosi seseorang terhadap topik tertentu seperti reaksi publik di media sosial terhadap tokoh politik atau isu (Wijaya et al., 2021). NLP bagian dari kecerdasan buatan, memproses teks dari berbagai sumber untuk mengidentifikasi emosi seperti sedih, gembira, atau marah (M. Furqan et al., 2023).

Penelitian terbaru NLP di Indonesia berfokus pada prediksi sentimen teks secara lebih efisien, termasuk pengembangan model BERT (*Bidirectional Encoder Representations from Transformers*). Beberapa model BERT khusus Bahasa Indonesia seperti IndoBERT dari IndoNLU dan IndoLEM telah banyak digunakan (Koto et al., 2020). IndoBERT dan IndoBERTweet, sebagai model *pre-trained* digunakan untuk mendeteksi emosi serta IndoBERT dilatih dengan algoritma BERT untuk Bahasa Indonesia, sementara IndoBERTweet dikembangkan dengan cakupan yang lebih luas (Wilie et al., 2019).

Penggunaan model bahasa IndoBERT yang telah dilatih dengan 4 miliar kata serta 250 juta kalimat memiliki arsitektur BERT namun menggunakan dataset pelatihan yang berbeda sehingga menjadikannya optimal dalam memahami konteks linguistik Bahasa Indonesia karena data memegang peranan penting dalam analisis sentimen terutama di platform media sosial seperti Instagram di mana komentar mencerminkan opini publik dan analisis sentimen memproses data komentar yang diperoleh melalui metode *web scraping* untuk mengumpulkan data secara efisien.

Web scraping mempermudah pengumpulan data secara otomatis dari situs web termasuk komentar yang diposting di media sosial Instagram namun

penggunaannya dalam konteks analisis sentimen Instagram memerlukan beberapa pertimbangan (Ratnasari, C. I & Khirana Dwicahyo, 2023) berdasarkan penelitian sebelumnya berbagai metode web scraping telah dikembangkan seperti *Xpath Selector* *JSON Parsing* *CSS Selector* *HTML Parsing* dan metode lainnya (Levi et al., 2020).

Prambors FM menyatakan bahwa terdapat 99,9 juta pengguna aktif bulanan Instagram di Indonesia sehingga menempatkan negara ini di posisi keempat setelah India Amerika Serikat dan Brasil sementara itu publikasi suara.com mencatat jumlah pengguna Instagram di Indonesia mencapai 99,15 juta orang yang mewakili 35,7% dari total penduduk Indonesia hal ini menjadikan Indonesia peringkat ketiga terbesar di dunia dalam jumlah pengguna media sosial yang berdampak pada nilai serta intensitas percakapan tentang isu-isu penting di berbagai platform media sosial (Toni A, 2022).

Instagram memudahkan manajemen dan komunikasi sekaligus memungkinkan pengguna terhubung dengan orang di seluruh dunia sehingga tidak heran jika platform ini sangat populer di kalangan pengguna gadget selain itu Instagram juga dimanfaatkan sebagai media komunikasi politik untuk menyampaikan informasi yang diperlukan dan penggunaannya dalam pilpres kini sudah umum dalam pemilihan umum (PEMILU) para kandidat memanfaatkan Instagram karena keunggulannya dalam menyebarkan informasi dengan cepat (Rismawati & Suryanef, 2021).

Sejalan dengan penjelasan tersebut fokus penelitian ini adalah analisis sentimen terhadap komentar pengguna Instagram dalam konteks pemilihan presiden di Indonesia dengan menggunakan model IndoBERT dan IndoBERTweet untuk memproses data teks berbahasa Indonesia yang dikumpulkan melalui metode *web scraping* fokus utama penelitian ini adalah mengungkap serta mengkategorikan pendapat pengguna Instagram guna memahami apakah tanggapan mereka terhadap kandidat bersifat *netral positif* atau *negatif* serta menyajikan pemahaman yang lebih mendalam mengenai dinamika *opini publik* di media sosial selama proses pemilihan umum.

1.2. Rumusan Masalah

Dari latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana performa model IndoBERT dalam mengklasifikasikan sentimen komentar Instagram terhadap capres 2024?
2. Bagaimana perbandingan akurasi dan efektivitas antara model IndoBERT dan IndoBERTweet dalam analisis sentimen komentar Instagram terhadap capres 2024?

1.3. Tujuan Penelitian

Penelitian ini memiliki tujuan yaitu :

1. Menganalisis performa model IndoBERT dalam mengklasifikasikan sentimen komentar Instagram terkait capres 2024.
2. Membandingkan akurasi dan efektivitas antara model IndoBERT dan IndoBERTweet dalam analisis sentimen komentar Instagram terkait capres 2024.

1.4. Batasan Masalah

Adapun batasan masalah dari penelitian ini mencakup:

1. Penelitian ini hanya pada penggunaan bahasa pemrograman *Python* sebagai bahasa pemrograman utama. *Python* dipilih karena memiliki banyak library dan *framework* yang mendukung proses *web scraping*, *preprocessing*, serta penerapan model *machine learning* seperti IndoBERT dan IndoBERTweet.
2. Dataset yang digunakan dalam penelitian ini diperoleh melalui *web scraping* dari platform media sosial Instagram @totalpolitikcom , @kompastv dan @detikcom. Batasan pada *scraping* adalah komentar yang berkaitan dengan calon presiden (capres) 2024, sehingga hanya data dari akun atau postingan terkait yang diambil. Hasil *scraping*

yang tidak *relevan*, seperti komentar spam atau tidak terkait dengan topik, akan dihilangkan melalui tahap *preprocessing*.

3. Penelitian ini hanya akan menggunakan model IndoBERT dan IndoBERTweet untuk melakukan analisis sentimen. Model-model lain tidak akan dibandingkan atau digunakan dalam penelitian ini.
4. Sentimen yang dianalisis dalam penelitian ini adalah tiga kategori utama, yaitu *netral*, *positif* dan *negati*. Komentar yang tidak dapat dikategorikan ke dalam salah satu dari tiga kelas tersebut akan diabaikan. Pengukuran *performa* model juga akan difokuskan pada akurasi dalam mengklasifikasikan komentar ke dalam kategori sentimen tersebut.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi:

1. Memberikan wawasan mendalam mengenai dinamika opini publik di media sosial selama pemilihan umum.
2. Memberikan bukti empiris tentang efektivitas model IndoBERT dan IndoBERTweet pada analisis sentimen berbahasa Indonesia.
3. Dapat dijadikan panduan praktis untuk peneliti lain yang ingin menggunakan metode *web scraping* untuk analisis data media sosial.
4. Membantu calon kandidat dan tim kampanye memahami sentimen publik dan menyesuaikan strategi komunikasi mereka.