



ITS
Institut
Teknologi
Sepuluh Nopember

KERJA PRAKTIK - EF234603

Studi Komparatif Skenario Fine-Tuning Model Transformer untuk Analisis Sentimen Teks Twitter Berbahasa Indonesia

Laboratorium Manajemen Cerdas Informasi
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota Surabaya,
Jawa Timur 60117

Periode: 1 Juli 2024 - 30 September 2024

Oleh:

Fadhl Akmal Madany 5025221028

Muhammad Rifqi Ma'ruf 5025221060

Pembimbing Jurusan

Shintami Chusnul Hidayati, S.Kom., M.Sc., Ph.D.

Pembimbing Lapangan

Ratih Nur Esti Anggraini, S.Kom., M.Sc., Ph.D.

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2024



KERJA PRAKTIK - EF234603
Studi Komparatif Skenario Fine-Tuning Model Transformer
untuk Analisis Sentimen Teks Twitter Berbahasa Indonesia

Laboratorium Manajemen Cerdas Informasi
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota Surabaya, Jawa
Timur 60117

Periode: 1 Juli 2024 - 30 September 2024

Oleh:

Fadhl Akmal Madany 5025221028

Muhammad Rifqi Ma'ruf 5025221060

Pembimbing Jurusan

Shintami Chusnul Hidayati, S.Kom., M.Sc., Ph.D.

Pembimbing Lapangan

Ratih Nur Esti Anggraini, S.Kom., M.Sc., Ph.D.

DEPARTEMEN TEKNIK INFORMATIKA

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya 2024

[Halaman ini sengaja dikosongkan]

DAFTAR ISI

DAFTAR ISI	4
DAFTAR GAMBAR	8
DAFTAR TABEL	10
LEMBAR PENGESAHAN	12
KATA PENGANTAR	17
BAB I	
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan	4
1.4. Manfaat	4
1.5. Lokasi dan Waktu Kerja Praktik	6
1.6. Metodologi Kerja Praktik	6
1.7. Sistematika Laporan	7
BAB II	
PROFIL PERUSAHAAN	11
2.1. Profil Laboratorium Manajemen Cerdas Informasi	11
2.2. Logo Perusahaan	11
2.3. Visi Misi Perusahaan	12
2.3.1. Visi Departemen Teknik Informatika	12
2.3.2. Misi Departemen Teknik Informatika	13
2.4. Struktur Organisasi	13
BAB III	
TINJAUAN PUSTAKA	15
3.1. Analisis Sentimen	15
3.2. Tokenisasi	15
3.3. Transfer Learning	16
	4

3.4. BERT	16
3.5. RoBERTa	18
3.6. IndoNLU	18
3.7. IndoBERT	19
3.8. IndoBERTweet	19
3.9. Discriminative Fine-Tuning	20
3.10. Gradual Layer Unfreezing	21
BAB IV	
IMPLEMENTASI SISTEM	23
4.1. Analisis Alur Sistem	23
4.2. Tahap Pemahaman Data	24
4.3. Tahap Pemrosesan Data	26
4.4. Konfigurasi Eksperimen dan Hyperparameter	28
4.5. Proses Pelatihan dan Evaluasi	29
BAB V	
PENGUJIAN DAN EVALUASI	33
5.1. Skenario 1: Baseline - FineTuning	33
5.2. Skenario 2: Gradual Freezing	34
5.3. Skenario 3: Differential Learning Rates	35
5.4. Evaluasi dan Pembahasan Hasil	36
5.4.1. Analisis Komparatif Kinerja Model	36
5.4.2. Efektivitas Skenario Fine-Tuning Standar	37
5.4.3. Gradual Unfreezing vs. Differential Learning Rates	37
5.4.4. Analisis Pengaruh Karakteristik Dataset terhadap Performa	38
BAB VI	
KESIMPULAN DAN SARAN	41
6.1. Kesimpulan	41

6.2. Saran	41
DAFTAR PUSTAKA	44
BIODATA PENULIS I	47
BIODATA PENULIS II	47

[Halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

Gambar 3.1. Kerangka Operasi <i>pre-training</i> dan <i>fine tuning</i>	15
Gambar 4.1. <i>flowchart</i> Analisis Sentimen Data Twitter.....	24

[Halaman ini sengaja dikosongkan]

DAFTAR TABEL

Tabel 4.1 Persebaran Dataset Sentimen BBM.....	16
Tabel 4.2 Persebaran Dataset Sentimen Pemilu.....	18
Tabel 4.3 Sampel Dataset Sentimen BBM.....	18
Tabel 4.4 Sampel Dataset Sentimen Pemilu.....	22
Tabel 4.5 Contoh Penerapan Teknik Pemrosesan Teks.....	28
Tabel 4.6 Konfigurasi Skenario 1: Fine-Tuning baseline.....	29
Tabel 4.7 Konfigurasi Skenario 2: Gradual Unfreezing.....	30
Tabel 4.8 Konfigurasi Skenario 3: Differential Learning Rates...	30
Tabel 5.1 Pengujian Dataset BBM Skenario 1.....	35
Tabel 5.2 Pengujian Dataset Pemilu Skenario 1.....	35
Tabel 5.3 Pengujian Dataset BBM & Pemilu Skenario 2.....	36
Tabel 5.4 Pengujian Dataset BBM & Pemilu Skenario 3.....	37

[Halaman ini sengaja dikosongkan]

**LEMBAR PENGESAHAN
KERJA PRAKTIK**

**Studi Komparatif Skenario Fine-Tuning Model
Transformer untuk Analisis Sentimen Teks Twitter
Berbahasa Indonesia**

Oleh:

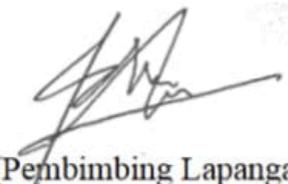
Fadhl Akmal Madany	5025221028
Muhammad Rifqi Ma'ruf	5025221060

Disetujui oleh Pembimbing Kerja Praktik:

1. Shintami Chusnul Hidayati,
S.Kom., M.Sc., Ph.D.
NIP. 1987202012004


(Pembimbing Departemen)

2. Ratih Nur Esti Anggraini,
S.Kom., M.Sc., Ph.D.
NIP. 198412102014042003


(Pembimbing Lapangan)

Studi Komparatif Skenario Fine-Tuning Model Transformer untuk Analisis Sentimen Teks Twitter Berbahasa Indonesia

Nama Mahasiswa : Fadhl Akmal Madany
NRP : 5025221028
Nama Mahasiswa : Muhammad Rifqi Ma'ruf
NRP : 5025221060
Departemen : Teknik Informatika FTEIC-ITS
Pembimbing Departemen : Shintami Chusnul Hidayati, S.Kom.,
M.Sc., Ph.D.
Pembimbing Lapangan : Ratih Nur Esti Anggraini, S.Kom.,
M.Sc., Ph.D.

ABSTRAK

Pertumbuhan konten digital berbahasa Indonesia yang pesat menuntut adanya sistem analisis sentimen yang akurat dan efisien. Model *transformer* berbasis *transfer learning* telah menunjukkan potensi besar, namun strategi *fine-tuning* yang paling efektif untuk konteks bahasa Indonesia masih memerlukan eksplorasi lebih lanjut. Penelitian ini melakukan studi komparatif untuk mengevaluasi tiga skenario *fine-tuning*—*Fine-Tuning Standar*, *Gradual Unfreezing*, dan *Differential Learning Rates*—pada tiga arsitektur model: IndoBERTbase, IndoBERTweet, dan RoBERTa. Pengujian dilakukan pada dua dataset dengan domain berbeda, yaitu ulasan aplikasi (Dataset BBM) dan komentar politik (Dataset Pemilu), dengan metrik evaluasi utama F1-Score. Hasil penelitian menunjukkan bahwa IndoBERTweet secara konsisten menjadi model dengan performa tertinggi di semua skenario, mencapai F1-Score puncak 0.9218

pada Dataset BBM dan 0.7431 pada Dataset Pemilu. Temuan kunci lainnya adalah strategi *Fine-Tuning* Standar dengan *learning rate* yang teroptimasi terbukti lebih unggul dibandingkan dua teknik lanjutan yang lebih kompleks. Penelitian ini menyimpulkan bahwa keselarasan domain antara data *pre-training* model dengan data tugas akhir merupakan faktor krusial untuk mencapai performa tinggi, dan sebuah metode yang lebih sederhana namun teroptimasi dengan baik dapat lebih efektif daripada strategi yang kompleks.

Kata Kunci : Analisis Sentimen, Transfer Learning, Fine-Tuning, BERT, IndoBERT, IndoBERTweet, RoBERTa, Pemrosesan Bahasa Alami, Bahasa Indonesia.

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat, karunia, serta petunjuk-Nya sehingga penulis dapat menyelesaikan laporan Kerja Praktik (KP). Penyusunan laporan ini merupakan salah satu syarat wajib untuk menyelesaikan program studi pada Departemen Teknik Informatika, Institut Teknologi Sepuluh Nopember.

Penulis menyadari bahwa dalam penyusunan laporan Kerja Praktik ini masih terdapat banyak kekurangan, baik dari segi metodologi penelitian maupun kedalaman analisis. Meskipun demikian, penulis berharap laporan ini dapat memberikan tambahan wawasan, serta menjadi referensi yang berguna bagi pembaca yang tertarik pada bidang Pemrosesan Bahasa Alami (Natural Language Processing) dan *Machine Learning*.

Pada kesempatan ini, penulis ingin mengucapkan rasa terima kasih yang tulus kepada semua pihak yang telah memberikan dukungan, bimbingan, dan semangat selama pelaksanaan hingga penyusunan laporan Kerja Praktik ini, antara lain:

1. Kedua orang tua penulis,
2. Ibu Ratih Nur Esti Anggraini, S.Kom., M.Sc., Ph.D. , selaku dosen pembimbing Kerja Praktik yang telah dengan sabar memberikan bimbingan yang berharga dan Shintami Chusnul Hidayati, S.Kom., M.Sc., Ph.D. selaku pembimbing jurusan.

Surabaya, 11 Juli 2025

Fadhl Akmal Madany dan Muhammad Rifqi Ma'ruf

[Halaman ini sengaja dikosongkan]

BAB I

PENDAHULUAN

1.1. Latar Belakang

Era digital kontemporer ditandai oleh ledakan volume data yang belum pernah terjadi sebelumnya, terutama dari interaksi pengguna di platform media sosial. Twitter, misalnya, telah berevolusi menjadi ruang diskursus publik global, di mana jutaan pengguna secara aktif menyampaikan opini, sentimen, dan tanggapan terhadap isu-isu aktual secara real-time (Albladi, Islam, & Seals, 2025). Di Indonesia, Twitter sering digunakan sebagai barometer opini publik dalam menanggapi isu-isu sosial-politik dan ekonomi, seperti pemilihan umum (pemilu) dan kenaikan harga Bahan Bakar Minyak (BBM).

Analisis data dari media sosial seperti Twitter memberikan peluang besar dalam memahami dinamika opini publik. Informasi ini sangat penting bagi pengambil kebijakan, akademisi, dan pelaku industri karena dapat digunakan sebagai dasar pengambilan keputusan yang lebih responsif (Liu, 2012). Analisis sentimen, sebagai bagian dari text mining, bertujuan mengidentifikasi serta mengevaluasi informasi subjektif dan keadaan emosional dari teks, dan menjadi salah satu pendekatan utama dalam menggali opini publik secara sistematis.

Namun demikian, menganalisis data teks dari Twitter—khususnya dalam Bahasa Indonesia—bukanlah hal yang mudah. Tantangan muncul dari sifat teks yang tidak terstruktur, informal, dan seringkali mengandung bahasa gaul, singkatan, kesalahan ketik, serta campuran bahasa (code-switching) antara Bahasa Indonesia dan Inggris. Ekspresi ironi dan sarkasme juga membuat interpretasi sentimen menjadi lebih kompleks (Cahyawijaya et al., 2021). Oleh karena itu, pendekatan

tradisional seperti metode berbasis leksikon seringkali tidak mampu menangani keragaman linguistik tersebut secara efektif.

Kemajuan teknologi dalam bidang Machine Learning dan Deep Learning telah memberikan solusi yang lebih andal. Akan tetapi, melatih model dari awal memerlukan dataset yang sangat besar dan sumber daya komputasi yang tidak selalu tersedia untuk Bahasa Indonesia. Transfer learning menjadi solusi yang menjanjikan karena memungkinkan model yang telah dilatih pada korpus besar untuk disesuaikan kembali melalui fine-tuning pada dataset yang lebih kecil dan spesifik (Devlin, Chang, Lee, & Toutanova, 2019).

Arsitektur Transformer dengan mekanisme self-attention menjadi fondasi dari model-model mutakhir seperti BERT (Bidirectional Encoder Representations from Transformers). Untuk Bahasa Indonesia, telah dikembangkan model seperti IndoBERT (Wilie et al., 2020) dan IndoBERTweet (Koto, Lau, & Baldwin, 2021), yang dilatih secara khusus menggunakan data berbahasa Indonesia dan Twitter. IndoBERT menawarkan pemahaman mendalam terhadap struktur sintaksis dan semantik bahasa Indonesia, sedangkan IndoBERTweet dirancang secara spesifik untuk menangani karakteristik informal dan khas dari teks Twitter Indonesia. Selain itu, model global seperti RoBERTa juga menunjukkan performa yang unggul dengan pendekatan pre-training yang lebih optimal (Liu et al., 2019).

Walaupun ketiga model ini menunjukkan potensi besar, performa mereka sangat bergantung pada proses fine-tuning yang optimal, terutama dalam hal pemilihan hyperparameter seperti learning rate, batch size, dan jumlah epoch. Penyesuaian hyperparameter dapat berdampak signifikan terhadap akurasi dan kemampuan generalisasi model terhadap data baru (Albladi et al.,

2025). Oleh karena itu, penelitian yang mengevaluasi baik performa model maupun pengaruh tuning hyperparameter menjadi sangat penting.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk melakukan analisis komparatif terhadap tiga model transfer learning—IndoBERT, IndoBERTweet, dan RoBERTa—untuk tugas analisis sentimen pada dua domain isu di Twitter Indonesia: isu politik (pemilu) dan sosial-ekonomi (kenaikan BBM). Penelitian ini juga akan mengevaluasi pengaruh dari hyperparameter tuning pada performa masing-masing model guna mengidentifikasi pendekatan terbaik yang robust dan efektif dalam konteks analisis sentimen berbahasa Indonesia.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, terdapat beberapa tantangan dan kebutuhan dalam analisis sentimen pada data Twitter berbahasa Indonesia yang memerlukan investigasi lebih lanjut. Oleh karena itu, permasalahan utama dalam penelitian ini dapat dirumuskan dalam bentuk pertanyaan-pertanyaan berikut:

- a. Model manakah di antara IndoBERTbase, IndoBERTweet, dan RoBERTa yang menunjukkan performa paling unggul untuk tugas analisis sentimen, dan sejauh mana keselarasan domain data pre-training mempengaruhi keunggulan tersebut?
- b. Apakah strategi fine-tuning lanjutan seperti Gradual Unfreezing dan Differential Learning Rates mampu memberikan peningkatan performa yang signifikan dibandingkan dengan strategi Fine-Tuning Standar yang hyperparameter-nya telah dioptimalkan?
- c. Bagaimana karakteristik dataset (seperti kompleksitas domain dan kualitas data)

mempengaruhi performa akhir dari setiap model dan strategi fine-tuning yang diuji?

1.3. Tujuan

Selaras dengan rumusan masalah di atas, penelitian ini memiliki beberapa tujuan utama sebagai berikut:

- a. Mengidentifikasi arsitektur model untuk analisis sentimen berbahasa Indonesia dengan menganalisis pengaruh keselarasan domain antara data pre-training dan data tugas akhir.
- b. Mengevaluasi secara komparatif efektivitas dari tiga strategi fine-tuning untuk menentukan apakah metode lanjutan memberikan keunggulan praktis di atas metode standar yang telah dioptimalkan dengan baik.
- c. Menganalisis dan memahami bagaimana perbedaan dalam kompleksitas dan kualitas data antara dua domain yang berbeda (kenaikan bahan bakar minyak dan komentar pemilu) secara signifikan mempengaruhi hasil performa model dan efektivitas strategi fine-tuning.

1.4. Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan manfaat baik secara akademis maupun praktis, antara lain:

- a. Manfaat Akademis:
 - Memberikan kontribusi empiris pada literatur Natural Language Processing (NLP) untuk Bahasa Indonesia, khususnya dengan menyediakan analisis perbandingan performa

head-to-head antara model IndoBERT, IndoBERTweet, dan Roberta pada domain yang spesifik dan relevan.

- Menyajikan benchmark performa yang dapat menjadi acuan bagi peneliti selanjutnya yang bekerja pada dataset atau tugas analisis sentimen sejenis di Indonesia.
- Memberikan wawasan mengenai pentingnya adaptasi model terhadap domain data (misalnya, data umum vs. data Twitter) dan efektivitas hyperparameter tuning dalam konteks transfer learning untuk bahasa non-Inggris.

b. Manfaat Praktis:

- Memberikan panduan praktis bagi para praktisi industri, analis data, dan ilmuwan sosial dalam memilih arsitektur model transfer learning yang paling efektif untuk kebutuhan analisis sentimen pada data media sosial di Indonesia.
- Menyediakan rekomendasi model dan konfigurasi optimal yang dapat diterapkan secara langsung untuk aplikasi nyata seperti riset pasar, pemantauan citra merek (brand monitoring), analisis kebijakan publik, dan pemetaan opini politik.
- Membantu organisasi dalam menghemat waktu dan sumber daya komputasi dengan memberikan titik awal yang kuat untuk pengembangan model analisis sentimen, tanpa harus melakukan eksperimen ekstensif dari awal.

1.5. Lokasi dan Waktu Kerja Praktik

Kerja praktik dilaksanakan dengan deskripsi sebagai berikut:

Lokasi : Dalam jaringan

Waktu : 1 Juli – 30 September 2024

Hari Kerja : Senin – Minggu

Jam Kerja : Fleksibel (minimal 15 jam per minggu)

1.6. Metodologi Kerja Praktik

a. Perumusan Masalah

Metodologi kerja praktik ini dirancang untuk memberikan alur kerja yang sistematis dan terstruktur dalam mencapai tujuan penelitian. Berdasarkan rumusan masalah yang telah diajukan, metodologi ini disusun melalui beberapa tahapan utama, mulai dari studi literatur hingga evaluasi model transfer learning. Pendekatan ini memastikan bahwa setiap langkah proses dijalankan secara efektif dan sesuai dengan tantangan yang ada, sehingga analisis yang dihasilkan akurat dan dapat diandalkan.

b. Studi Literatur

Setelah masalah dirumuskan, dilakukan studi literatur mendalam. Tahap ini melibatkan eksplorasi sumber ilmiah untuk mengumpulkan informasi relevan mengenai Analisis Sentimen, Transfer Learning, arsitektur Transformer (BERT, RoBERTa, IndoBERT, IndoBERTweet), dan teknik Hyperparameter Tuning.

c. Implementasi Sistem

Berdasarkan pemahaman yang diperoleh dari studi literatur, kerangka kerja komputasi dibangun untuk mendukung eksperimen. Tahap ini mencakup analisis alur sistem, pemahaman dan pemrosesan data, konfigurasi eksperimen dan *hyperparameter*, serta proses pelatihan model.

d. Pengujian dan Evaluasi

Setelah model diimplementasikan, dilakukan pengujian dan evaluasi untuk mengukur kinerjanya. Pengujian ini bertujuan menemukan *learning rate* optimal dan mengevaluasi komparatif skenario *fine-tuning* yang berbeda, seringkali dengan mengukur F1-Score dari proses *cross-validation*.

e. Kesimpulan dan Saran

Pengujian dan evaluasi yang dilakukan telah memenuhi syarat yang diinginkan, dan berjalan dengan baik dan lancar dalam konteks penelitian analisis sentimen.

1.7. Sistematika Laporan

a. Bab I Pendahuluan

Bab ini menyajikan pengantar dari keseluruhan penelitian. Bagian ini mencakup latar belakang yang menguraikan konteks dan urgensi masalah, rumusan masalah yang menjadi fokus penelitian, tujuan yang ingin dicapai, serta manfaat yang diharapkan baik secara akademis maupun praktis. Bab ini juga menjelaskan metodologi kerja praktik secara umum dan diakhiri dengan sistematika laporan untuk memberikan gambaran struktur penulisan.

b. Bab II Profil Perusahaan

Bab ini berisi profil Laboratorium Manajemen Cerdasi Informasi selaku klien kerja praktik, baik dari segi lokasi, visi, misi, struktur organisasi, dan aspek lainnya.

c. Bab III Tinjauan Pustaka

Bab ini berisi landasan teori yang menjadi dasar dari penelitian. Bagian ini mengulas berbagai hasil peninjauan pustaka dari jurnal, prosiding konferensi, buku, dan sumber terpercaya lainnya. Topik yang dibahas mencakup teori fundamental mengenai Analisis Sentimen, Transfer Learning, serta arsitektur model-model relevan seperti BERT, RoBERTa, IndoBERT, dan IndoBERTtweet.

d. Bab IV Metode Penelitian

Bab ini menjelaskan secara rinci langkah-langkah teknis yang dilakukan dalam penelitian. Bagian ini diawali dengan diagram alir kerja praktik untuk memberikan gambaran keseluruhan proses. Selanjutnya, dijabarkan setiap tahapan secara mendalam, mulai dari deskripsi dan persiapan dataset, proses preprocessing data, hingga perancangan skenario eksperimen dan metode evaluasi model.

e. Bab V Analisis dan Pembahasan

Bab ini menyajikan hasil dari seluruh eksperimen yang telah dilakukan. Bagian ini menampilkan data performa dari setiap skenario model dalam bentuk tabel dan visualisasi. Selain menyajikan hasil, bab ini juga berisi analisis dan pembahasan mendalam mengenai temuan tersebut, termasuk perbandingan kinerja antar model dan dampak dari optimasi hyperparameter.

f. Bab VII Kesimpulan dan Saran

Bab ini merupakan penutup dari laporan kerja praktik. Bagian ini berisi penarikan kesimpulan yang secara langsung menjawab rumusan masalah dan tujuan penelitian. Selain itu, pada bab ini juga diberikan saran-saran konstruktif untuk pengembangan atau penelitian lebih lanjut di masa mendatang berdasarkan temuan dan keterbatasan dari penelitian yang telah dilakukan.

[Halaman ini sengaja dikosongkan]

BAB II

PROFIL PERUSAHAAN

2.1. Profil Laboratorium Manajemen Cerdas Informasi

Laboratorium Manajemen Cerdas Informasi adalah salah satu laboratorium riset dibawah Departemen Teknik Informatika Institut Teknologi Sepuluh Nopember (ITS) yang berfokus pada pengembangan bidang analisis, sintesis, dan evaluasi proses bisnis dan sistem informasi pada sistem *Enterprise*, implementasikan rekayasa pengetahuan, melakukan audit dan tata kelola Teknologi Informasi, serta merancang solusi basis data terdistribusi dan *Big Data*. Beberapa mata kuliah pendukungnya antara lain: Sistem *Enterprise*, Rekayasa Pengetahuan, Audit Sistem, Tata Kelola Teknologi Informasi, Basis Data Terdistribusi, dan *Big Data*. Laboratorium ini juga menjadi basis untuk berbagai penelitian di tingkat pascasarjana.

2.2. Logo Perusahaan

Gambar 2.1 menampilkan logo Laboratorium Manajemen Cerdas Informasi (MCI) Teknik Informatika ITS.



Gambar 2.1 Logo Lab MCI

2.3. Visi Misi Perusahaan

2.3.1. Visi Departemen Teknik Informatika

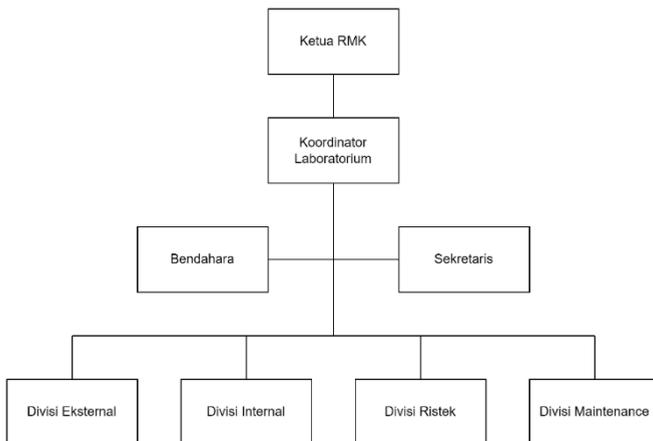
Sejalan dengan visi ITS yaitu menjadi perguruan tinggi dengan reputasi internasional dalam ilmu pengetahuan, teknologi, dan seni, terutama yang menunjang industri dan kelautan yang berwawasan lingkungan, maka visi Departemen Informatika adalah menjadi inovator bidang informatika yang unggul di tingkat nasional dengan reputasi internasional, serta berperan aktif dalam upaya memajukan dan mensejahterakan bangsa.

Visi Departemen Teknik Informatika (DTIF) berfokus untuk menjadi lembaga pendidikan yang unggul dan inovatif-inventif dalam informatika cerdas, mendukung transformasi digital dan berkontribusi kepada masyarakat dengan reputasi internasional.

2.3.2. Misi Departemen Teknik Informatika

1. Menyelenggarakan proses pembelajaran yang berkualitas, dan memenuhi standar nasional maupun internasional,
2. Melaksanakan penelitian yang inovatif, bermutu, dan bermanfaat,
3. Meningkatkan pemanfaatan teknologi informasi dan komunikasi untuk masyarakat,
4. Menjalin kemitraan dengan berbagai lembaga, baik di dalam maupun di luar negeri.

2.4. Struktur Organisasi



Gambar 2.2 Struktur Organisasi MCI

Struktur organisasi lab MCI terdiri dari ketua RMK, koordinator laboratorium, bendahara, sekretaris, divisi eksternal, divisi internal, divisi ristik, dan divisi *maintenance* sebagaimana yang ditampilkan pada Gambar 2.2.

[Halaman ini sengaja dikosongkan]

BAB III

TINJAUAN PUSTAKA

3.1. Analisis Sentimen

Analisis Sentimen, juga disebut sebagai *Opinion Mining*, adalah bidang studi yang berfokus pada teknik komputasional untuk menganalisis teks untuk mengidentifikasi dan memahami opini, sentimen, emosi, dan evaluasi yang diungkapkan di dalamnya. Bidang ini sering kali berpusat pada konten yang dibuat pengguna di media sosial. Analisis dapat dilakukan pada berbagai tingkatan, termasuk klasifikasi tingkat dokumen, analisis tingkat kalimat, dan tingkat berbasis aspek. Tujuan utamanya adalah untuk secara formal mendefinisikan dan mengekstrak komponen opini, seperti siapa yang mengungkapkannya (pemegang opini), tentang apa itu (entitas dan aspeknya), sentimen itu sendiri, dan kapan opini itu dibagikan (Saberri & Saad, 2017).

3.2. Tokenisasi

Tokenisasi adalah teknik dalam *Natural Language Processing* (NLP) yang melibatkan pembagian teks masukan menjadi unit-unit yang lebih kecil dan bermakna yang disebut "token". Token-token ini merepresentasikan kemunculan kata-kata individual dalam sebuah teks dan kemudian dimasukkan ke dalam tahap-tahap NLP berikutnya seperti analisis morfologi, penandaan kelas kata, dan penguraian. Tugas tambahan tokenisasi sering kali mencakup identifikasi batas kalimat, karena langkah-langkah pemrosesan selanjutnya biasanya beroperasi pada kalimat-kalimat individual. Meskipun

bukan langkah pertama mutlak dalam abstraksi (perbedaan penyusunan huruf biasanya disaring terlebih dahulu), tokenisasi sering digunakan karena mengubah teks mentah menjadi format yang sesuai untuk analisis linguistik, sehingga memungkinkan tugas-tugas NLP untuk dilakukan secara efektif (Grefenstette, 1999).

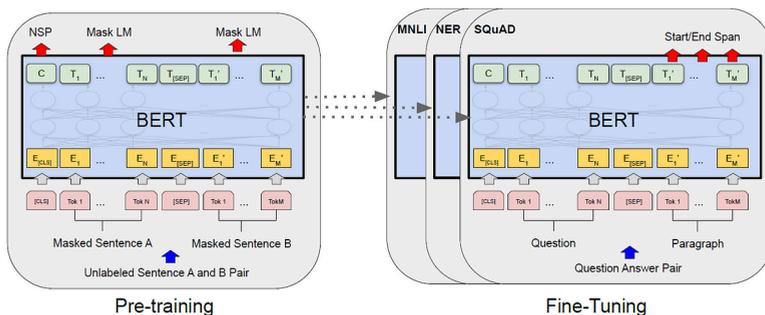
3.3. Transfer Learning

Transfer Learning adalah metodologi di mana model pertama-tama memperoleh pengetahuan umum dengan pelatihan pada dataset besar dan kemudian menerapkan atau mentransfer pengetahuan tersebut ke tugas yang berbeda, yang seringkali lebih spesifik. Metode ini telah menjadi teknik yang efektif dalam bidang seperti *computer vision*, di mana model dilatih terlebih dahulu pada basis data ImageNet yang besar dan kemudian disempurnakan untuk tujuan lain (Devlin et al., 2019).

3.4. BERT

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model representasi bahasa yang dirancang untuk melatih representasi dua arah dari teks yang tidak berlabel. Tidak seperti model sebelumnya yang menggunakan model bahasa searah (memeriksa teks dari kiri ke kanan atau kanan ke kiri), BERT dirancang untuk secara bersamaan mengkondisikan konteks kiri dan kanan di semua lapisan. Kerangka BERT terdiri dari dua fase operasi: *pre-training* dan *fine-tuning*, sebuah penerapan dari konsep *transfer learning*, dimana model pertama-tama memperoleh pengetahuan dari data dalam skala besar, kemudian mentransfer pengetahuan tersebut untuk disesuaikan dengan tugas-tugas baru yang lebih

spesifik. Dalam fase *pre-training*, model dilatih pada data tak berlabel menggunakan dua *unsupervised tasks*, yang memungkinkannya mengembangkan pemahaman komprehensif tentang struktur dan hubungan bahasa. Setelah itu, fase *fine-tuning* mengadaptasi model BERT yang telah dilatih sebelumnya untuk tugas-tugas *Natural Language Processing* (NLP) yang lebih spesifik. Hal ini dicapai dengan menginisialisasi model baru dengan *pre-trained parameter*, menambahkan satu lapisan keluaran, lalu *fine-tuning* seluruh parameter (*end-to-end*) menggunakan data berlabel untuk suatu tugas tersebut (Devlin et al., 2019).



Gambar 3. 1 Kerangka operasi *pre-training* dan *fine-tuning* untuk BERT (Devlin et al., 2019).

Tahap *pre-training* dibedakan oleh dua tujuan: *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Tujuan MLM memungkinkan model untuk mempelajari konteks dua arah. Dalam tugas ini, 15% dari token input dalam suatu urutan ditutupi secara acak, dan tujuan model adalah untuk memprediksi token yang ditutupi ini berdasarkan konteks yang tidak ditutupi di sekitarnya. Tujuan kedua, NSP, melatih model untuk memahami hubungan antara kalimat. Untuk tugas ini,

model disajikan dengan dua kalimat, A dan B, dan harus memprediksi apakah kalimat B adalah kalimat sebenarnya yang mengikuti A dalam teks asli atau hanya kalimat acak dari korpus. Kombinasi kedua tugas ini memungkinkan BERT untuk mempelajari representasi tingkat token dan tingkat kalimat secara bersamaan.

3.5. RoBERTa

RoBERTa, singkatan dari *Robustly Optimized BERT Pretraining Approach*, adalah model bahasa yang dibangun dan mengoptimalkan model BERT (*Bidirectional Encoder Representations from Transformers*). Dikembangkan oleh para peneliti di Facebook AI dan University of Washington, RoBERTa merupakan hasil studi replikasi BERT yang menemukan bahwa model asli kurang terlatih secara signifikan. Modifikasi utama dalam RoBERTa meliputi pelatihan model untuk durasi yang lebih lama dengan batch yang lebih besar pada lebih banyak data, menghapus tujuan *Next Sentence Prediction* (NSP), pelatihan pada urutan yang lebih panjang, dan secara dinamis mengubah pola *masking* yang diterapkan pada data pelatihan. Dengan penyesuaian ini, RoBERTa mampu mencapai hasil *state-of-the-art* pada beberapa *benchmark* pemrosesan bahasa alami, termasuk GLUE, RACE, dan SQuAD, melampaui kinerja model BERT sebelumnya (Liu et al., 2019).

3.6. IndoNLU

IndoNLU adalah kerangka *benchmark* pertama untuk bahasa Indonesia yang diusulkan karena kelangkaan sumber daya *Natural Language Processing* (NLP) untuk bahasa Indonesia. Kerangka ini mencakup dua belas tugas

evaluasi yang beragam, mulai dari analisis sentimen hingga *named entity recognition*, untuk mengukur pemahaman bahasa alami. Dataset untuk tugas-tugas ini berasal dari berbagai domain dan gaya untuk memastikan keragaman tugas (Wilie et al., 2020).

3.7. IndoBERT

Bersamaan dengan tolok ukur IndoNLU, diperkenalkan model bahasa IndoBERT yang berbasis BERT dan dilatih secara khusus untuk diuji pada IndoNLU. Model ini dilatih menggunakan Indo4B, sebuah dataset masif berisi sekitar empat miliar kata dari berbagai sumber teks bahasa Indonesia seperti berita, media sosial, dan blog. Ketika kinerjanya diuji menggunakan IndoNLU, model IndoBERT menunjukkan performa yang sangat kompetitif. Secara umum, IndoBERT berhasil mencapai skor rata-rata tertinggi pada tugas-tugas klasifikasi dan mengungguli model-model multibahasa pada 8 dari 12 tugas yang diuji (Wilie et al., 2020).

3.8. IndoBERTweet

IndoBERTweet adalah model bahasa *pre-trained* skala besar pertama untuk Twitter Indonesia, yang dikembangkan dengan memperluas model BERT monolingual Indonesia, IndoBERT. Model ini dibuat untuk mengatasi tantangan dalam memperoleh representasi kontekstual berkualitas tinggi untuk domain media sosial. Model ini dilatih pada dataset 26 juta tweet Indonesia yang dikumpulkan antara Desember 2019 dan Desember 2020. Fokus utama pengembangan IndoBERTweet adalah menemukan metode yang efisien untuk mengadaptasi model ke kosakata baru yang

ditemukan di Twitter. IndoBERTweet menginisialisasi *embedding* untuk kata-kata baru yang spesifik domain dengan merata-ratakan *embedding* subkata dari model IndoBERT. Teknik ini tidak hanya membuat proses *pre-training* lima kali lebih cepat daripada pelatihan dari awal tetapi juga menghasilkan kinerja yang lebih baik pada tujuh tugas berbasis Twitter yang berbeda. Tugas-tugas ini meliputi analisis sentimen, klasifikasi emosi, deteksi *hate speech*, dan *named entity recognition* (Koto et al., 2021).

3.9. Discriminative Fine-Tuning

Discriminative Fine-Tuning adalah metode yang memungkinkan setiap lapisan model disetel dengan *learning rate* yang berbeda (*Discriminative Learning Rates*), berdasarkan prinsip bahwa lapisan yang berbeda menangkap berbagai jenis informasi. Dalam update Stochastic Gradient Descent (SGD) standar, satu *learning rate* berlaku untuk semua parameter model. *Discriminative Fine-Tuning* memodifikasi hal tersebut dengan membagi parameter model, θ , ke dalam satu set untuk setiap lapisan, dari $\{\theta^1, \dots, \theta^L\}$, dimana L adalah jumlah total lapisan. Setiap lapisan θ^l kemudian diberi *learning rate*, η^l , dan SGD update dilakukan sesuai untuk setiap lapisan (Howard & Ruder, 2018).

Howard dan Ruder (2018) mengusulkan pendekatan empiris, dimana *Learning Rate* untuk lapisan akhir, η^L , ditentukan terlebih dahulu. Laju pembelajaran untuk lapisan sebelumnya kemudian ditetapkan relatif terhadap laju lapisan terakhir menggunakan formula

$$\eta^{l-1} = \eta^l / 2.6$$

Persamaan 3.1 Discriminative Learning Rate

3.10. Gradual Layer Unfreezing

Gradual Layer Unfreezing adalah teknik *fine-tuning* untuk mencegah *catastrophic forgetting*, dimana model kehilangan informasi berguna yang diperoleh selama *pre-training*. Alih-alih menyempurnakan semua lapisan secara bersamaan, metode ini melibatkan *unfreezing* model secara progresif mulai dari lapisan terakhir, yang dianggap memiliki pengetahuan umum yang paling sedikit.

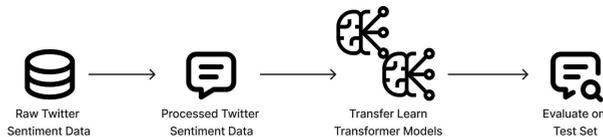
Proses ini dimulai dengan *unfreezing* hanya lapisan terakhir dan menyempurnakannya selama satu epoch. Selanjutnya, dilakukan *unfreezing* pada lapisan bawah berikutnya, dan proses ini diulang. Proses tersebut berlanjut hingga telah terjadi *unfreezing* pada semua lapisan model dan dilatih hingga konvergensi. Pendekatan ini dicatat mirip dengan '*chain-thaw*', tetapi alih-alih melatih satu lapisan pada satu waktu, metode ini memperluas rangkaian lapisan yang 'tercairkan' dengan setiap langkah (Howard & Ruder, 2018).

[Halaman ini sengaja dikosongkan]

BAB IV IMPLEMENTASI SISTEM

4.1. Analisis Alur Sistem

Alur sistem dalam penelitian ini, seperti yang ditunjukkan pada gambar 4.1, dimulai dengan Tahap Pemahaman Data, di mana dua dataset sentimen mentah (BBM dan Pemilu) dengan karakteristik dan distribusi sentimen dianalisis. Data ini kemudian akan diproses yang melibatkan normalisasi, pembersihan ekspresi reguler, penggantian slang, tokenisasi, penghapusan *stopword*, dan *stemming* untuk menstandarisasi teks. Selanjutnya, model transformer *pre-trained* (IndoBERT, RoBERTa, IndoBERTtweet) di-*fine-tune* menggunakan teknik *transfer learning*, *gradual unfreezing*, dan *discriminative learning rates* dengan *hyperparameter* yang telah dioptimalkan. Proses pelatihan melibatkan pembagian data, iterasi *epoch*, *forward pass*, perhitungan *loss*, *backpropagation*, dan optimisasi. Terakhir, sistem masuk ke tahap Tahap Evaluasi, di mana model diuji menggunakan data yang belum pernah dilihat untuk mengukur kinerja dengan metrik standar (akurasi, presisi, recall, f1-score) dan *Confusion Matrix*, memastikan validitas generalisasi model.



Gambar 4.1 *flowchart* Analisis Sentimen Data Twitter

4.2. Tahap Pemahaman Data

Pada penelitian ini, akan digunakan dua dataset, yakni dataset sentimen mengenai kenaikan harga Bahan Bakar Minyak (BBM) dan dataset sentimen mengenai Pemilu. Dataset pertama, yang berfokus pada BBM, terdiri dari 7.529 baris data, sedangkan dataset kedua, yang membahas Pemilu, memiliki total 42.692 baris data. Keduanya berisi teks berbahasa Indonesia yang bersifat nonformal, seperti opini dan cuitan dari media sosial, yang telah diklasifikasikan ke dalam sentimen positif, negatif, atau netral.

Sebagai tambahan, kedua dataset ini menunjukkan perbedaan fundamental dalam distribusi sentimennya. Dataset BBM cenderung tidak seimbang (*imbalanced*), dengan sentimen 'Netral' mendominasi secara signifikan (65,01%), seperti yang ditunjukkan pada Tabel 4.1. Sebaliknya, dataset Pemilu memiliki distribusi yang jauh lebih seimbang di antara ketiga kategori sentimen (Positif 29,53%, Netral 36,06%, dan Negatif 34,39%), detailnya dapat dilihat pada Tabel 4.2. Perbedaan ini, ditambah dengan karakteristik teks yang mengandung bahasa non-baku, singkatan, dan emoticon (contoh dapat dilihat pada Tabel 4.3 dan Tabel 4.4), menjadikan kedua dataset ini kasus uji yang baik. Dataset ini dapat dimanfaatkan untuk mengevaluasi performa model analisis sentimen dalam menangani data dengan distribusi yang berbeda sekaligus menguji kemampuannya dalam memahami teks yang kompleks dan beragam dari media sosial.

Tabel 4.1 Persebaran Dataset Sentimen BBM

Sentimen	Jumlah	Persentase
Positif	507	6,73%
Netral	4894	65,01%

Negatif	2128	28,26%
Total	7529	100%

Tabel 4.2 Persebaran Dataset Sentimen Pemilu

Sentimen	Jumlah	Persentase
Positif	12610	29,53%
Netral	15398	36,06%
Negatif	14684	34,39%
Total	42692	100%

Tabel 4.3 Sampel Dataset Sentimen BBM

Sentimen	Text
Positif	@KakekHalal Mereka yg menghina pak Jokowi adalah orang 2 yang belum kenal pak Jokowi Pak Jokowi merakyat Di era beliau BBM GG pernah naik GG ada korupsi GG ada merangkap jabatan Pencipta mobil Esemka.... Hebat
Netral	Ini daftar harga BBM Pertamina di seluruh Indonesia https://t.co/C68YbHzLst
Negatif	@Nast_1205 @bachrum_achmadi Kau yg tolol. Rakyat disuruh byr pajak trus gajikan manusia2 tak berguna. Kalo berguna

	tdk mungkin BBM meroket goblok...
--	-----------------------------------

Tabel 4.4 Sampel Dataset Sentimen Pemilu

Sentimen	Text
Positif	Pemerintah tetap berkomitmen untuk menjalankan pemilu sesuai tahapan
Netral	kita dah buat yang terbaik tapi masih tetap rasa kecewa
Negatif	izinkan aku mengupload ini

4.3. Tahap Pemrosesan Data

Pada tahap ini, data teks dipersiapkan terlebih dahulu untuk analisis sentimen. Intuisi utamanya adalah membersihkan dan menstandarisasi teks, yang sering kali *noisy* dan informal, terutama karena bersumber dari media sosial. Pemrosesan memastikan bahwa model dapat berfokus pada konten teks yang relevan, yang menghasilkan performa dan generalisasi yang lebih baik.

Prosesnya dimulai dengan normalisasi teks yang melibatkan pengubahan semua teks menjadi huruf kecil untuk memperlakukan kata-kata seperti "BBM" dan "bbm" sebagai identik. Ekspresi reguler kemudian digunakan untuk menghapus artefak media sosial umum seperti penyebutan pengguna (@username), tagar (#), URL, dan indikator retweet ("rt"). Langkah ini menghilangkan *noise*, yang memungkinkan model untuk berkonsentrasi pada bagian substantif teks.

Setelah normalisasi, akan dilakukan transformasi teks. Kata-kata slang diganti dengan padanan formalnya dalam bahasa Indonesia. Teks tersebut kemudian ditokenisasi, atau dipecah menjadi kata-kata individual. Setelah tokenisasi, kata-kata umum dengan nilai semantik yang kecil, yang dikenal sebagai stopword (misalnya, "dan", "di", "yang"), dihilangkan menggunakan *library* Sastrawi. Akhirnya, setiap kata di-stem ke bentuk akarnya (misalnya, "kenaikan" menjadi "naik"), yang membantu mengurangi ukuran kosakata dan mengelompokkan kata-kata terkait menjadi satu. Contoh penerapan teknik pemrosesan pada salah satu data teks dapat dilihat pada Tabel 4.5.

Tabel 4.5 Contoh Penerapan Teknik Pemrosesan Teks

Teknik Pemrosesan	Text
Teks Asli	@OposisiCerdas @venasantika Gaji besar bukannya kerja bener malah bikin gaduh, gw ga setuju!
Huruf Kecil	@oposisi cerdas @venasantika gaji besar bukannya kerja bener malah bikin gaduh, gw ga setuju!
Pembersihan Ekspresi Reguler	gaji besar bukannya kerja bener malah bikin gaduh gw ga setuju
Penggantian Slang	gaji besar bukannya kerja benar malah bikin gaduh saya tidak setuju
Tokenisasi	['gaji', 'besar', 'bukannya', 'kerja', 'benar', 'malah', 'bikin', 'gaduh', 'saya', 'tidak', 'setuju']

Penghapusan Stopword	['gaji', 'besar', 'kerja', 'benar', 'bikin', 'gaduh', 'setuju']
----------------------	---

4.4. Konfigurasi Eksperimen dan Hyperparameter

Setiap skenario dijalankan pada model IndoBERTbase, IndoBERTtweet, dan RoBERTa dengan konfigurasi dan set hyperparameter spesifik yang dirangkum dalam tabel-tabel berikut. Optimizer yang digunakan di semua skenario adalah AdamW, yang merupakan standar untuk melatih model *transformer*.

Tabel 4.6 Konfigurasi Skenario 1: Fine-Tuning baseline

Hyperparameter	Nilai
Learning Rate	5e-05, 3e-05, 2e-05, 1e-05
Batch Size	16
Epoch	4
Cross-Validation	Stratified 3-Fold

Tabel 4.7 Konfigurasi Skenario 2: Gradual Unfreezing

Hyperparameter	Tahap 1	Tahap 2	Tahap 3
Epoch	1	2	2
Batch Size	1e-3	1e-5	2e-6

Lapisan Aktif	Classifier	Classifier + 4 Lapisan Atas	Seluruh Model
---------------	------------	--------------------------------------	------------------

Tabel 4.8 Konfigurasi Skenario 3: Differential Learning Rates

Grup Lapisan	Learning Rate
Classifier & Pooler	5e-05
4 Lapisan Atas	2e-05
4 Lapisan Tengah	1e-05
Lapisan Bawah & Embeddings	5e-06

4.5. Proses Pelatihan dan Evaluasi

Pada tahap pelatihan ini, akan dilakukan *fine-tuning* pada setiap model *pre-trained* agar dapat mengenali pola sentimen yang spesifik pada dataset BBM dan Pemilu. Proses *fine-tuning* ini akan diuji melalui tiga skenario yang berbeda: *Fine-Tuning Baseline* (Skenario 1), *Gradual Unfreezing* (Skenario 2), dan *Differential Learning Rates* (Skenario 3). Untuk membantu proses ini, setiap dataset yang telah melalui tahap pra-pemrosesan dibagi menjadi tiga bagian: 80% untuk data pelatihan (training set), 10% untuk data validasi (validation set), dan 10% untuk data pengujian (test set). Pelatihan dilakukan secara iteratif dalam beberapa epoch, di mana pada setiap epoch, model memproses seluruh data

pelatihan dalam batch kecil untuk mengoptimalkan penggunaan memori.

Siklus pelatihan untuk setiap batch dimulai dengan forward pass, di mana data input yang telah di-tokenisasi dimasukkan ke dalam model untuk menghasilkan prediksi sentimen. Prediksi ini kemudian dibandingkan dengan label sentimen sebenarnya (ground truth) untuk menghitung tingkat kesalahan, atau loss. Nilai loss ini menjadi dasar bagi proses backpropagation, yang menyebarkan informasi kesalahan kembali ke seluruh jaringan untuk menghitung gradien dari setiap parameter. Terakhir, optimizer menggunakan gradien ini untuk memperbarui bobot model, dengan tujuan meminimalkan loss secara bertahap.

Pseudocode 3.1 Proses Pelatihan per Epoch (training_epoch)

```
FUNCTION training_epoch(model, dataloader,
optimizer, scaler):
    SET model ke mode training
    INITIALIZE total_loss, all_predictions,
all_labels

    FOR each batch in dataloader:
        RESET optimizer gradients
        MOVE data ke perangkat (GPU/CPU)

        // Forward Pass dengan presisi campuran
untuk efisiensi
        WITH autocast():
            outputs = model(input_ids,
attention_mask, labels)
            loss = outputs.loss
        END WITH

        // Backward Pass dan update bobot
SCALE loss menggunakan scaler
PERFORM backpropagation pada loss yang
```

```

telah di-scale
    UPDATE bobot model menggunakan optimizer
dan scaler

    STORE prediksi dan label dari batch
    ADD loss ke total_loss
END FOR

    CALCULATE metrik (F1, dll.) dari
all_predictions dan all_labels
    RETURN rata-rata loss, metrik performa
END FUNCTION

```

Pseudocode 3.2 Proses Evaluasi per Epoch (evaluation epoch)

```

FUNCTION evaluation_epoch(model, dataloader):
    SET model ke mode evaluasi
    INITIALIZE total_loss, all_predictions, all_labels

    // Nonaktifkan komputasi gradien untuk efisiensi
memori dan kecepatan
    WITH no_grad():
        FOR each batch in dataloader:
            MOVE data ke perangkat

            // Forward Pass
            outputs = model(input_ids, attention_mask, labels)
            loss = outputs.loss

            STORE prediksi dan label dari batch
            ADD loss ke total_loss
        END FOR
    END WITH

    CALCULATE metrik (F1, dll.) dari all_predictions dan
all_labels

```

```
RETURN rata-rata loss, metrik performa  
END FUNCTION
```

[Halaman ini sengaja dikosongkan]

BAB V PENGUJIAN DAN EVALUASI

5.1. Skenario 1: Baseline - FineTuning

Menggunakan Tabel 4.6 didapatkan hasil skenario pengujian pada dataset BBM dan Pemilu berturut turut dapat dilihat pada tabel di 5.1 dan 5.2

Tabel 5.1 Pengujian Dataset BBM Skenario 1

Learning Rates	Model					
	IndoBERTbase		IndoBERTtweet		RoBERTa	
	F1 Score	Val loss	F1 Score	Val loss	F1 Score	Val loss
1e-05	0.8990	0.2714	0.9144	0.2214	0.8702	0.3180
2e-05	0.9113	0.2329	0.9185	0.2656	0.9057	0.2636
3e-05	0.9156	0.2328	0.9218	0.3620	0.9075	0.2753
5e-05	0.9182	0.2612	0.9204	0.5197	0.9090	0.3259

Tabel 5.2 F1 Score Pengujian Dataset Pemilu Skenario 1

Learning Rates	Model					
	IndoBERTbase		IndoBERTtweet		RoBERTa	
	F1 Score	Val loss	F1 Score	Val loss	F1 Score	Val loss
1e-05	0.6985	0.6735	0.7268	0.6466	0.6901	0.6840

2e-05	0.7173	0.6423	0.7396	0.6912	0.7080	0.6621
3e-05	0.7267	0.6377	0.7431	0.7906	0.7151	0.6697
5e-05	0.7348	0.6306	0.7394	0.8773	0.7211	0.6639

Dari hasil di atas, IndoBERTweet secara konsisten menunjukkan performa tertinggi pada kedua dataset, dengan F1-Score puncak 0.9218 pada Dataset BBM (LR 3e-05) dan 0.7431 pada Dataset Pemilu (LR 3e-05). Hal ini mengindikasikan bahwa arsitektur IndoBERTweet yang telah dilatih pada data media sosial berbahasa Indonesia sangat cocok untuk tugas analisis sentimen pada kedua domain tersebut. Hasil ini akan menjadi *baseline* kuat untuk perbandingan dengan skenario-skenario berikutnya.

5.2. Skenario 2: Gradual Freezing

Menggunakan konfigurasi Tabel 4.7 didapatkan Hasil skenario pengujian pada dataset BBM dan Pemilu berturut turut dapat dilihat pada tabel berikut.

Tabel 5.3 Pengujian Dataset BBM & Pemilu Skenario 2

Dataset	Model					
	IndoBERTbase		IndoBERTweet		RoBERTa	
	F1 Score	Val loss	F1 Score	Val loss	F1 Score	Val loss
BBM	0.8916	0.3084	0.9042	0.2836	0.4890	0.8368

Pemilu	0.6721	0.7153	0.6921	0.6875	0.5474	0.9126
--------	--------	--------	--------	--------	--------	--------

Teknik Gradual Unfreezing menunjukkan peningkatan performa bertahap yang baik untuk model IndoBERTbase dan IndoBERTtweet. Namun, performa akhir dari kedua model ini tidak berhasil melampaui hasil terbaik dari Skenario 1. Hal yang paling menonjol adalah performa RoBERTa yang sangat rendah. F1-Score yang dihasilkan (0.4890 untuk BBM dan 0.5474 untuk Pemilu) jauh di bawah ekspektasi. Analisis lebih lanjut menunjukkan bahwa hal ini kemungkinan besar disebabkan oleh perbedaan arsitektural, di mana model RoBERTa yang digunakan tidak memiliki layer pooler dan memiliki classifier head yang lebih "tipis". Akibatnya, pada Tahap 1 di mana seluruh base model dibekukan, classifier head tidak mendapatkan sinyal yang cukup baik untuk belajar secara efektif, yang berdampak pada performa keseluruhan.

5.3. Skenario 3: Differential Learning Rates

Menggunakan Tabel 4.8 didapatkan hasil skenario pengujian pada dataset BBM dan Pemilu berturut turut dapat dilihat pada tabel di 5.1 dan 5.2

Tabel 5.4 Pengujian Dataset BBM & Pemilu Skenario 3

Dataset	Model					
	IndoBERTbase		IndoBERTtweet		RoBERTa	
	F1 Score	Val loss	F1 Score	Val loss	F1 Score	Val loss

BBM	0.9122	0.2532	0.9159	0.2862	0.8530	0.4211
Pemilu	0.7075	0.6599	0.7260	0.6600	0.6148	0.8327

Teknik Differential Learning Rates menghasilkan performa yang sangat kompetitif, terutama pada Dataset BBM. IndoBERTweet kembali menjadi yang terbaik dengan F1-Score 0.9159. Perlu dicatat bahwa RoBERTa menunjukkan pemulihan performa yang signifikan pada Dataset BBM (0.8530) dibandingkan dengan kegagalannya di Skenario 2, yang menandakan bahwa teknik ini lebih cocok untuk arsitektur RoBERTa. Namun, pada Dataset Pemilu, performa semua model dengan teknik ini masih berada di bawah hasil Skenario 1.

5.4. Evaluasi dan Pembahasan Hasil

5.4.1. Analisis Komparatif Kinerja Model

Salah satu temuan paling konsisten dari seluruh rangkaian pengujian adalah superioritas model IndoBERTweet. Pada kedua dataset (BBM dan Pemilu) dan di semua skenario pengujian, IndoBERTweet secara konsisten menghasilkan F1-Score tertinggi. Keunggulan ini dapat diatribusikan secara langsung pada keselarasan domain (antara data pre-training model dengan data yang digunakan dalam penelitian ini).

IndoBERTweet secara spesifik dilatih menggunakan korpus besar data dari media sosial Indonesia, termasuk Twitter. Karakteristik teks pada media sosial—seperti penggunaan bahasa informal, slang, singkatan, ekspresi emotif, dan kalimat pendek—sangat mirip dengan karakteristik data pada dataset Komentar BBM dan Komentar Pemilu. Oleh karena itu, representasi

fitur yang telah dipelajari oleh IndoBERTtweet selama pre-training sudah sangat relevan untuk tugas analisis sentimen pada domain ini, sehingga proses fine-tuning menjadi lebih efektif dan efisien. Sebaliknya, IndoBERTbase yang lebih banyak dilatih pada data formal (berita, Wikipedia) dan RoBERTa generik, meskipun berkinerja baik, memerlukan adaptasi yang lebih signifikan.

5.4.2. Efektivitas Skenario *Fine-Tuning* Standar

Temuan signifikan menunjukkan bahwa Skenario 1 (Fine-Tuning Standar) dengan *learning rate* yang teroptimasi, secara tak terduga mengungguli Skenario 2 (Gradual Unfreezing) dan Skenario 3 (Differential LR). Fenomena ini dapat dijelaskan melalui dua hipotesis: pertama, kemungkinan dataset yang memadai mencegah *catastrophic forgetting* parah pada model IndoBERTtweet, sehingga fine-tuning standar sudah cukup untuk adaptasi parameter. Kedua dan yang lebih krusial, keberhasilan Skenario 1 kemungkinan besar disebabkan oleh optimasi *hyperparameter* utama (*learning rate*) yang menyeluruh. Sementara itu, teknik-teknik yang lebih kompleks pada skenario lain kurang melalui proses *tuning* yang ekstensif, menegaskan prinsip bahwa metode yang lebih sederhana namun teroptimasi dengan baik seringkali dapat melampaui pendekatan yang lebih kompleks dengan konfigurasi standar.

5.4.3. Gradual Unfreezing vs. Differential Learning Rates

Saat membandingkan kedua teknik lanjutan, Skenario 3 (*Differential LR*) secara konsisten terbukti jauh lebih unggul dibandingkan Skenario 2 (*Gradual Unfreezing*),

terutama dengan kegagalan signifikan Skenario 2 pada model RoBERTa. Gradual Unfreezing merupakan pendekatan yang 'kaku'; pada tahap awal, pembekuan seluruh base model memaksa classifier head yang baru diinisialisasi untuk belajar dari representasi fitur yang statis. Ini menjadi masalah jika classifier head 'tipis' atau memiliki parameter terbatas (seperti RoBERTa tanpa pooler), mengakibatkan kapasitas belajar yang tidak memadai dan performa awal yang buruk. Sebaliknya, Differential LR menawarkan pendekatan yang lebih 'fleksibel' dan 'holistik', memungkinkan seluruh model untuk belajar secara simultan dengan kecepatan yang bervariasi. Hal ini menjaga aliran gradien yang stabil di seluruh jaringan, memungkinkan adaptasi yang lebih robust bagi classifier head dan lapisan atas, menjadikannya lebih efektif untuk arsitektur modern seperti RoBERTa.

5.4.4. Analisis Pengaruh Karakteristik Dataset terhadap Performa

Perbedaan performa yang signifikan terlihat antara Dataset BBM (F1-Score > 0.90) dan Dataset Pemilu (F1-Score ~ 0.74), yang dapat diatribusikan pada karakteristik data yang berbeda. Teks terkait pemilu secara inheren lebih kompleks karena sarat dengan sarkasme, ironi, dan ambiguitas, menuntut pemahaman kontekstual yang mendalam, berbeda dengan sentimen ulasan aplikasi yang lebih lugas.

Selain itu, kualitas Dataset Pemilu teridentifikasi kurang konsisten, dengan adanya noise, teks tidak relevan, serta inkonsistensi pelabelan yang disebabkan oleh sifat subjektif topik politik. Tantangan-tantangan ini secara langsung tercermin pada nilai validation loss yang jauh lebih tinggi pada Dataset Pemilu di semua skenario

(misalnya, 0.6385 vs. 0.2699 pada Skenario 3). Loss yang tinggi ini mengindikasikan bahwa model kesulitan menemukan fungsi pemetaan yang pasti dari teks ke label akibat ambiguitas dan noise dalam data.

[Halaman ini sengaja dikosongkan]

BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Berdasarkan hasil studi komparatif skenario fine-tuning model Transformer untuk analisis sentimen teks Twitter berbahasa Indonesia, dapat disimpulkan beberapa poin sebagai berikut:

- a. Model IndoBERTweet menunjukkan performa paling unggul untuk tugas analisis sentimen pada kedua dataset yang diuji, mencapai F1-Score 0.7712 pada Dataset BBM dan 0.7431 pada Dataset Pemilu.
- b. Strategi Fine-Tuning Standar dengan learning rate yang teroptimasi terbukti lebih unggul dibandingkan dua teknik lanjutan yang lebih kompleks yaitu Gradual Unfreezing dan Differential Learning Rates.
- c. Karakteristik dataset, seperti kompleksitas domain dan kualitas data, sangat memengaruhi performa model dan efektivitas strategi fine-tuning.

6.2. Saran

Untuk pengembangan lebih lanjut di masa mendatang, disarankan beberapa hal berikut:

- a. Melakukan eksplorasi lebih lanjut terhadap metode optimasi hyperparameter yang lebih canggih di luar optimasi learning rate standar untuk strategi Gradual Unfreezing dan Differential Learning Rates untuk melihat apakah performa mereka dapat ditingkatkan lebih jauh.
- b. Menguji model dan skenario fine-tuning pada dataset sentimen berbahasa Indonesia dari domain yang berbeda (misalnya, ulasan produk, berita) untuk memvalidasi generalisasi temuan.

- c. Mempertimbangkan penggunaan teknik data augmentation untuk dataset yang tidak seimbang (seperti Dataset BBM) guna mengatasi masalah imbalanced class dan potensi peningkatan performa model.
- d. Menganalisis lebih dalam jenis kesalahan (error analysis) yang dilakukan oleh model pada sentimen yang salah diklasifikasikan untuk mengidentifikasi area perbaikan spesifik, terutama pada ekspresi ironi atau sarkasme yang kompleks dalam bahasa Indonesia.
- e. Menyelidiki dampak dari ukuran model yang berbeda (misalnya, base vs. large) pada arsitektur yang sama untuk analisis sentimen bahasa Indonesia.

[Halaman ini sengaja dikosongkan]

DAFTAR PUSTAKA

- Abbladi, A., Islam, M., & Seals, C. D. (2025). Sentiment analysis of Twitter data using NLP models: A comprehensive review. *IEEE Access*.
- Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., Ruder, S., Lim, Z. Y., Bahar, S., Khodra, M. L., Purwarianti, A., & Fung, P. (2021). IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation. In *Proceedings of EMNLP 2021* (pp. 8875–8898).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (2), 4171-4186. 10.18653/v1/N19-1423
- Grefenstette, G. (1999). Tokenization. In *Syntactic wordclass tagging* (pp. 117-133). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-9273-4_9
- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *arXiv preprint arXiv:1801.06146*, (5). <https://arxiv.org/pdf/1801.06146>
- Koto, F., Lau, J. H., & Baldwin, T. (2021). IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. *arXiv preprint arXiv:2109.04607*. <https://arxiv.org/pdf/2109.04607>
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019).

- Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
<https://arxiv.org/pdf/1907.11692>
- Saberi, B., & Saad, S. (2017). Sentiment Analysis or Opinion Mining: A Review. *Int. J. Adv. Sci. Eng. Inf. Technol*, 7(5), 1660-1666.
- Tala, F. (2003). *A study of stemming effects on information retrieval in Bahasa Indonesia*.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *arXiv preprint arXiv:2009.05387*.
<https://arxiv.org/pdf/2009.05387>

[Halaman ini sengaja dikosongkan]

BIODATA PENULIS I

Nama : Fadhl Akmal Madany
Tempat, Tanggal Lahir : Surabaya, 28 Januari 2004
Jenis Kelamin : Laki-Laki
Telepon : +6288231887320
Email : akmal.madany1@gmail.com

AKADEMIS

Kuliah : Departemen Teknik Informatika –
FTEIC , ITS
Angkatan : 2022
Semester : 6 (Enam)

BIODATA PENULIS II

Nama : Muhammad Rifqi Ma'ruf
Tempat, Tanggal Lahir : Sidoarjo, 08 Oktober 2003
Jenis Kelamin : Laki-laki
Telepon : +628977452963
Email : mrifqim.id@gmail.com

AKADEMIS

Kuliah : Departemen Teknik Informatika –
FTEIC , ITS
Angkatan : 2022
Semester : 6 (Enam)