

**ANALISIS SENTIMEN BERBASIS ASPEK  
PADA APLIKASI PELAYANAN MASYARAKAT  
MENGUNAKAN INDOBERT**

**LAPORAN TUGAS AKHIR**

Laporan ini disusun untuk memenuhi salah satu syarat memperoleh Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



**Disusun Oleh:**

**Muhammad Rafli**

**32602100086**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG**

**2025**

***FINAL PROJECT***

***Aspect-Based Sentiment Analysis on Public Service  
Applications using IndoBERT***

*proposed to complete the requirement to obtain a bachelor's degree (S-1)  
at Informatics Engineering Departement of Industrial Technology Faculty  
Sultan Agung Islamic University*



**Arranged By:**

**MUHAMMAD RAFLI**

**32602100086**

***MAJORING OF INFORMATICS ENGINEERING  
FACULTY OF INDUSTRIAL TECHNOLOGY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG***

**LEMBAR PENGESAHAN  
TUGAS AKHIR**

**ANALISIS SENTIMEN BERBASIS ASPEK PADA APLIKASI  
PELAYANAN MASYARAKAT MENGGUNAKAN INDOBERT**

**Muhammad Rafli  
NIM 32602100086**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir  
Program Studi Teknik Informatika  
Universitas Islam Sultan Agung  
Pada tanggal : 30-12-2025

**TIM PENGUJI UJIAN SARJANA :**

**Mustafa, S.T., M.M., M.Kom**  
NIK. 210610040  
(Ketua Penguji)

07-01-2026

**Andi Riansyah, S.T., M. Kom**  
NIK. 210616053  
(Anggota Penguji)

07-01-2026

**Badie'ah, S.T., M.Kom**  
NIK. 210615044  
(Pembimbing)

21-01-2026

Semarang, 20-01-2026

Mengetahui,

Kaprodi Teknik Informatika  
Universitas Islam Sultan Agung

**Moch. Taufik, S.T., M.IT**  
NIK. 2106040

## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Muhammad Rafli

NIM : 32602100086

Judul Tugas Akhir : ANALISIS SENTIMEN BERBASIS ASPEK PADA  
APLIKASI PELAYANAN MASYARAKAT  
MENGUNAKAN INDOBERT

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

UNISSULA  
جامعة سلطان أبوبوع الإسلامية

Semarang, 21-01-2026

Yang Menyatakan,



Muhammad Rafli

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Rafli

NIM : 32602100086

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :  
ANALISIS SENTIMEN BERBASIS ASPEK PADA APLIKASI PELAYANAN  
MASYARAKAT MENGGUNAKAN INDOBERT

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Scmarang, 21-01-2026

Yang menyatakan,



Muhammad Rafli

## KATA PENGANTAR

Dengan mengucapkan rasa syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan Rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Laporan Proposal Tugas Akhir dengan judul “Analisis Sentimen Berbasis Aspek Pada Aplikasi Pelayanan Masyarakat Menggunakan IndoBERT” ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan sebagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Dr. Gunarto, S.H., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, ST, MT., IPU., ASEAN Eng
3. Dosen Pembimbing penulis Bu Badie'ah, S.T., M.Kom yang banyak memberikan banyak nasehat dan saran .
4. Orang tua dan keluarga penulis yang telah menjadi *support system* dan mengizinkan untuk menyelesaikan laporan ini,
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan rendah hati,, penulis menyadari bahwa laporan masih memiliki banyak kekurangan dalam kualitas dan isi. Karena itu , penulis mengharapkan kritik dan saran yang membangun untuk penyempurnaan dimasa depan

Semarang, 08 Januari 2025



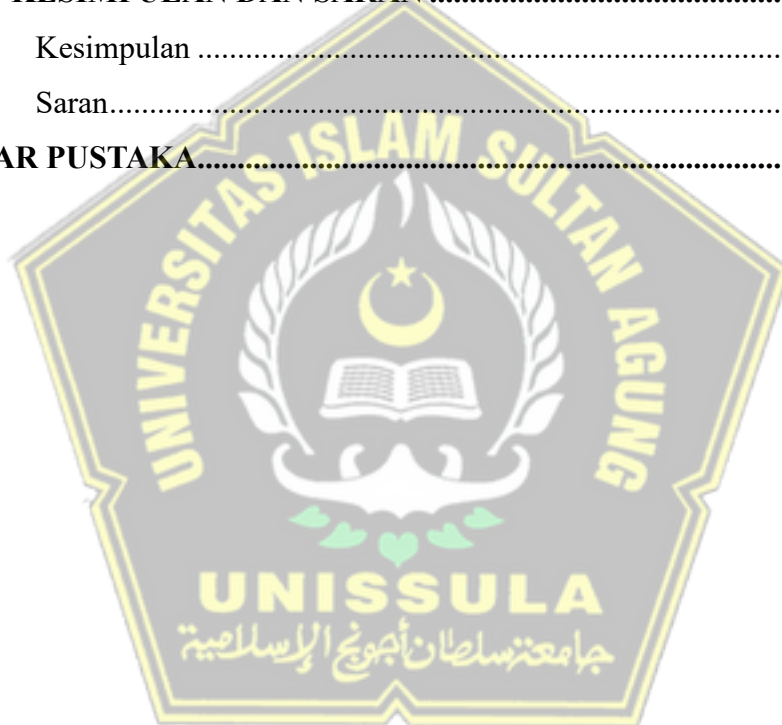
Muhammad Rafli

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN TUGAS AKHIR .....</b>	<b>iii</b>
<b>SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....</b>	<b>iv</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR .....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>ABSTRAK .....</b>	<b>xii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	2
1.3 Pembatasan Masalah .....	3
1.4 Tujuan Tugas Akhir.....	3
1.5 Manfaat .....	3
1.6 Sistematikan Penulisan .....	4
<b>BAB II TINJUAN PUSTAKA DAN DASAR TEORI.....</b>	<b>5</b>
2.1 Tinjauan Pustaka.....	5
2.2 DASAR TEORI .....	7
2.2.1 Analisis Sentimen .....	7
2.2.2 <i>Aspect-Based Sentiment Analysis (ABSA)</i> .....	7
2.2.3 Aspek Penelitian.....	7
2.2.4 Aplikasi Pelayanan Masyarakat .....	10
2.2.5 <i>Natural Language Processing (NLP)</i> .....	11
2.2.6 <i>Deep Learning</i> .....	12
2.2.7 <i>Transformers</i> .....	13

2.2.8	BERT.....	16
2.2.9	IndoBert.....	18
2.2.10	<i>Confusion Matrix</i> .....	19
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>19</b>
3.1	Metodologi Penelitian .....	19
3.2	Studi Literatur .....	20
3.3	Pengumpulan Data .....	20
3.4	Pra-Pemrosesan Data .....	24
3.4.1	<i>Text preprocessing</i> .....	25
3.4.2	<i>Labeling Data</i> .....	27
3.5	<i>Token classification</i> .....	31
3.6	<i>Fine-tuning</i> IndoBERT.....	35
3.7	<i>Evaluasi Model</i> .....	36
3.8	Analisis Hasil .....	36
3.9	Implementasi Sistem.....	37
3.9.1	Perancangan <i>interface</i> .....	37
3.9.2	Tahapan Kerja Sistem .....	38
<b>BAB IV HASIL DAN ANALISIS PENELITIAN.....</b>		<b>31</b>
4.1	Persiapan Data.....	31
4.2	Pra-Pemrosesan Data .....	33
4.2.1	<i>Text Preprocessing</i> .....	33
4.2.2	<i>Labelling data</i> .....	37
4.2.3	<i>Balancing</i> .....	40
4.3	<i>Token Classification</i> .....	42
4.4	<i>Fine-tuning</i> IndoBERT.....	46
4.4.1	<i>Spliting Data</i> .....	46
4.4.2	<i>Paramater</i> .....	46
4.5	Evaluasi Model.....	47

4.5.1	Evaluasi Model Mobile JKN.....	47
4.5.2	Evaluasi Model M-Paspor.....	52
4.5.3	Evaluasi Model SIGNAL.....	56
4.6	Analisis Performa Model .....	60
4.7	Hasil Implementasi Antarmuka.....	65
4.7.1	Hasil Perancangan Antarmuka .....	65
4.8	Pembahasan.....	70
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>59</b>
5.1	Kesimpulan .....	59
5.2	Saran.....	60
<b>DAFTAR PUSTAKA.....</b>		<b>61</b>



## DAFTAR GAMBAR

Gambar 2. 1	Arsitektur model <i>transformers</i> (Vaswani dkk., 2017).....	13
Gambar 2. 2	<i>Scaled Dot-Product Attention dan Multi-head attention</i> (Mohiuddin dkk., 2023).....	15
Gambar 2. 3	<i>Pre-training dan fine-tuning</i> (Devlin dkk., 2019).....	16
Gambar 3. 1	Alur metodologi penelitian.....	19
Gambar 3. 2	Alur pengumpulan data.....	22
Gambar 3. 3	Alur text preprocessing.....	25
Gambar 3. 4	Login label studio.....	28
Gambar 3. 5	Membuat <i>project</i> label studio.....	28
Gambar 3. 6	<i>Import Dataset</i> .....	29
Gambar 3. 7	Konfigurasi Label Studio.....	29
Gambar 3. 8	Pelabelan teks.....	30
Gambar 3. 9	Alur <i>token classification</i> .....	32
Gambar 3. 10	Halaman utama.....	37
Gambar 3. 11	Tahapan kerja pembuatan sistem.....	38
Gambar 4. 1	Grafik <i>Train Loss dan Validation Loss Mobile JKN</i> .....	50
Gambar 4. 2	Grafik <i>training dan validation loss M Paspor</i> .....	55
Gambar 4. 3	Grafik <i>training dan validation loss SIGNAL</i> .....	59
Gambar 4. 5	Tampilan <i>Interface</i> .....	66
Gambar 4. 6	Tampilan hasil deteksi multiaspek.....	66
Gambar 4. 7	Tampilan hasil deteksi <i>single</i> aspek pada model JKN.....	67
Gambar 4. 8	Tampilan hasil deteksi multiaspek pada model M-Paspor.....	68
Gambar 4. 9	Tampilan hasil deteksi <i>single</i> aspek pada model M-Paspor.....	68
Gambar 4. 10	Tampilan hasil deteksi multi aspek pada model SIGNAL.....	69
Gambar 4. 11	Tampilan hasil deteksi <i>single</i> aspek pada model SIGNAL.....	69

## DAFTAR TABEL

Tabel 3. 1 <i>Parameter</i> pengambilan data .....	22
Tabel 3. 2 <i>Case folding</i> .....	26
Tabel 3. 3 <i>Cleaning text</i> .....	26
Tabel 3. 4 <i>Normalization</i> .....	27
Tabel 3. 5 Labeling data .....	30
Tabel 3. 6 Tokenisasi teks.....	34
Tabel 3. 7 label BIO .....	34
Tabel 3. 8 Parameter <i>fine-tuning</i> .....	35
Tabel 4. 1 Hasil <i>Web Scraping</i> .....	32
Tabel 4. 2 <i>Case Folding</i> .....	33
Tabel 4. 3 Data <i>Sebelum dan Sesudah Cleaning pada Ulasan Aplikasi</i> .....	34
Tabel 4. 4 <i>Normalization</i> .....	36
Tabel 4. 5 Tabel Hasil Anotasi Data.....	37
Tabel 4. 6 Distribusi aspek dan sentimen Mobile JKN .....	39
Tabel 4. 7 Distribusi aspek dan sentimen M-Paspor .....	39
Tabel 4. 8 Distribusi aspek dan sentimen SIGNAL .....	39
Tabel 4. 9 Jumlah data Mobile JKN sebelum dan setelah <i>balancing</i> .....	41
Tabel 4. 10 Jumlah data M-Paspor sebelum dan setelah <i>balancing</i> .....	41
Tabel 4. 11 Jumlah data SIGNAL sebelum dan setelah <i>balancing</i> .....	41
Tabel 4. 12 Contoh Token Classification pada Data Ulasan .....	42
Tabel 4. 13 Spliting data peraplikasi .....	46
Tabel 4. 14 <i>Hyperparamater</i> .....	46
Tabel 4. 15 Matrik evaluasi dataset <i>training</i> Mobile JKN .....	47
Tabel 4. 16 Matrik evaluasi dataset <i>validation</i> Mobile JKN.....	48
Tabel 4. 17 Matrik evaluasi dataset <i>testing</i> Mobile JKN .....	49
Tabel 4. 18 Matrik evaluasi M Paspor .....	52
Tabel 4. 19 Matrik evaluasi data <i>validation</i> M Paspor.....	53
Tabel 4. 20 Matriks Evaluasi data <i>testing</i> M Paspor .....	54

Tabel 4. 21 Matrik evaluasi dataset <i>training</i> SIGNAL .....	56
Tabel 4. 22 Matrik Evaluasi data <i>validation</i> SIGNAL.....	57
Tabel 4. 23 Matrik Evaluasi data <i>testing</i> SIGNAL .....	58
Tabel 4. 24 Perbandingan accuracy model Mobile JKN, M-Paspor, SIGNAL.....	61
Tabel 4. 25 Perbandingan F1-Score per aspek pada Data <i>Testing</i> .....	62
Tabel 4. 26 Perbandingan <i>Precicion</i> dan Karakteristik Model .....	62
Tabel 4. 27 Perbandingan Test Loss dan Karakteristik Model.....	63
Tabel 4. 28 Perbandingan <i>Macro Average</i> dan <i>Weighted Average</i> dataset <i>Testing</i>	63



## ABSTRAK

Ulasan pengguna pada aplikasi layanan publik digital seperti Mobile JKN, M-Paspor, dan SIGNAL menyimpan wawasan penting bagi peningkatan kualitas layanan. Namun, analisis sentimen konvensional belum mampu mengidentifikasi aspek spesifik yang menjadi sumber kepuasan maupun keluhan pengguna. Penelitian ini mengembangkan sistem *Aspect-Based Sentiment Analysis* (ABSA) berbasis IndoBERT dengan pendekatan token classification. Sistem dirancang untuk mengekstraksi tiga aspek utama, yaitu *Usability*, *Reliability*, dan *Efficiency*, sekaligus mengklasifikasikan kategori sentimen. Hasil pengujian menunjukkan kinerja yang tinggi dengan akurasi *testing* sebesar 97% untuk Mobile JKN, 98% untuk M-Paspor, dan 98% untuk SIGNAL. Analisis lebih lanjut mengungkap bahwa aspek *Efficiency* secara konsisten menjadi yang paling mudah diklasifikasi dengan F1-Score di atas 0.94, sementara aspek *Reliability* menjadi tantangan utama dengan pola kesulitan berbeda pada setiap dataset. Temuan ini menunjukkan bahwa sistem yang dibangun efektif dalam memberikan wawasan terperinci guna mendukung peningkatan berkelanjutan layanan publik digital.

Keywords: Analisis Sentimen Berbasis Aspek, IndoBERT, *Token Classification*, Layanan Publik, Pemrosesan Bahasa Alami

## ABSTRACT

*User reviews of digital public service applications such as Mobile JKN, M-Paspor, and SIGNAL provide valuable insights for improving service quality. However, conventional sentiment analysis has not been able to identify specific aspects that drive user satisfaction or complaints. This study develops an Aspect-Based Sentiment Analysis (ABSA) system based on IndoBERT using a token classification approach. The system is designed to extract three key aspects, namely Usability, Reliability, and Efficiency, while simultaneously classifying sentiment categories. Experimental results demonstrate strong performance with testing accuracy of 97% for Mobile JKN, 98% for M-Paspor, and 98% for SIGNAL. Further analysis reveals that the Efficiency aspect is consistently the easiest to classify with F1-Score above 0.94, while the Reliability aspect poses the main challenge with different difficulty patterns across datasets. These findings indicate that the proposed system is effective in providing detailed insights to support the continuous improvement of digital public services.*

Keywords: *Aspect-Based Sentiment Analysis, IndoBERT, Token Classification, Public Services, Natural Language Processing*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Evolusi teknologi digital dan komunikasi telah mendorong perubahan mendasar yang mengubah cara masyarakat mengakses beragam layanan, termasuk layanan publik. Menanggapi perubahan ini pemerintah Indonesia menghadirkan sejumlah aplikasi digital seperti Mobile JKN, M-Paspor, dan Samsat Digital Nasional (SIGNAL) untuk meningkatkan efisiensi dan efektivitas. Aplikasi-aplikasi ini memberikan kemudahan nyata, mulai dari pelayanan kesehatan, pengurusan dokumen keimigrasian, hingga pembayaran pajak kendaraan, yang secara signifikan menghemat waktu dan sumber daya masyarakat.

Tingginya tingkat adopsi platform ini berbanding lurus dengan meluasnya penetrasi internet di Indonesia yang tercatat bahwa Indonesia telah menyentuh angka 69,21% pada tahun 2023 (Badan Pusat Statistik, 2024). Angka ini tidak hanya mengindikasikan bahwa platform digital telah menjadi jalur interaksi utama antara warga dan pemerintah, tetapi juga berimplikasi pada melimpahnya volume umpan balik pengguna dalam bentuk ulasan digital. Akibatnya ekspektasi publik terhadap mutu, keandalan, dan kecepatan respons layanan digital terus meningkat, menjadikan evaluasi kualitas layanan sebagai suatu keharusan.

Salah satu pendekatan umum untuk mengevaluasi kepuasan pengguna adalah melalui analisis sentimen konvensional. Namun, metode ini memiliki keterbatasan signifikan karena hanya mengklasifikasikan ulasan ke dalam tiga kategori besar: positif, negatif, dan netral. Analisis ini gagal menangkap aspek spesifik dari layanan yang menjadi sumber kepuasan atau keluhan, sehingga wawasan yang dihasilkan kurang mendalam dan sulit ditindaklanjuti (Putri dan Cahyono, 2024).

Untuk mengatasi kelemahan tersebut *Aspect-Based Sentiment Analysis* (ABSA) hadir dengan pendekatan yang jauh lebih mendalam. Alih-alih menghasilkan sentimen konvensional, ABSA mampu mengurai ulasan pengguna menjadi aspek-aspek fundamental seperti kemudahan registrasi, kecepatan server, dan kejelasan antarmuka, lalu menilai sentimen pada setiap aspek tersebut.

Kemampuan terperinci ini memungkinkan analisis yang tajam, bahkan dapat membedakan pujian pada satu fitur dan kritik pada fitur lainnya dalam ulasan yang sama. Implementasi ABSA yang efektif tentu membutuhkan model Pemrosesan Bahasa Alami yang andal. Dalam konteks Bahasa Indonesia, model *transformer* seperti IndoBERT menjadi pilihan utama karena telah terbukti memiliki pemahaman semantik yang mendalam dan mampu mencapai akurasi hingga 98% dalam tugas analisis sentimen menggunakan dataset data IndoNLU (Merdiansah dan Ali Ridha, 2024).

Berdasarkan uraian tersebut, penelitian ini bertujuan mengimplementasikan model IndoBERT untuk analisis sentimen berbasis aspek pada tiga aplikasi pelayanan masyarakat yang telah ditentukan Mobile JKN, M-Paspor, dan Samsat Digital Nasional (SIGNAL). Pemilihan ketiga aplikasi tersebut tidak dilakukan secara acak, melainkan didasarkan pada pertimbangan representatif terhadap beragam sektor layanan publik yang bersifat kritical dan berdampak langsung pada masyarakat, yaitu layanan kesehatan, keimigrasian, dan administrasi kendaraan bermotor. Penelitian ini diharapkan dapat menyediakan kerangka evaluasi yang memberikan wawasan mendalam dan terperinci bagi penyedia layanan publik dalam rangka peningkatan kualitas layanan secara berkelanjutan.

## 1.2 Perumusan Masalah

Perumusan masalah pada tugas akhir ini adalah:

1. Bagaimana membangun program *Aspect-Based Sentiment Analysis* (ABSA) menggunakan arsitektur model IndoBERT
2. Bagaimana performa model IndoBERT dalam melakukan klasifikasi sentimen berbasis aspek pada ulasan aplikasi layanan publik.
3. Aspek layanan apa yang paling dominan menerima sentimen positif dan negatif berdasarkan aspek yang telah ditentukan *Usability*, *Reliability*, *Efficiency* dari pengguna pada ulasan aplikasi Mobile JKN, M-Paspor, dan SIGNAL berdasarkan hasil analisis model yang dikembangkan.

### 1.3 Pembatasan Masalah

Berikut adalah pembatasan masalah dari tugas akhir ini:

1. Penelitian ini hanya menganalisis ulasan *Mobile* JKN, M-Paspor, dan SIGNAL di *Google Play Store*.
2. Analisis hanya dilakukan pada ulasan yang ditulis berbahasa Indonesia.
3. klasifikasi sentiment hanya ulasan positif dan negatif.
4. Penelitian hanya berfokus pada aspek-aspek *Usability* (kemudahan penggunaan), *Reliability* (Keandalan), dan *Efficiency* yang merujuk pada atribut kualitas perangkat lunak dalam standar ISO/IEC 25010.

### 1.4 Tujuan Tugas Akhir

Adapun tujuan dari tugas akhir ini antara lain:

1. Mengembangkan sistem berbasis *Aspect-Based Sentiment Analysis* (ABSA) pada aplikasi pelayanan masyarakat dengan mengimplementasikan arsitektur model IndoBERT untuk dapat mendeteksi aspek dan polaritas sentimen dalam satu kalimat secara bersamaan.
2. Mengidentifikasi performa model IndoBERT dalam melakukan klasifikasi sentimen berbasis aspek pada ulasan aplikasi layanan publik .
3. Mengidentifikasi dan menganalisis aspek layanan (*Usability, Reliability, Efficiency*) yang dominan menerima sentimen positif dan negatif pada setiap aplikasi berdasarkan hasil dari model ABSA yang dikembangkan.

### 1.5 Manfaat

Berikut merupakan manfaat dari tugas akhir ini:

1. Memberikan landasan evaluasi bagi penyedia layanan publik yang berbasis data secara spesifik dan dapat ditindak lanjuti. Informasi mengenai aspek *Usability, Reliability, dan Efficiency* yang mendapat keluhan maupun apresiasi dari pengguna sehingga dapat dijadikan fokus utama dalam upaya perbaikan dan pengembangan aplikasi ke depannya.
2. Memberikan kontribusi di bidang *Natural Language Processing* (NLP) Bahasa Indonesia, khususnya dalam penggunaan model *transformer*

(IndoBERT) dalam domain analisis sentimen yang berfokus pada aspek tertentu.

## 1.6 Sistematisan Penulisan

- BAB I : PENDAHULUAN**  
 Bab I menyajikan pendahuluan yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta uraian tentang sistematika penulisan.
- BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI**  
 BAB II menguraikan landasan teori dan kajian pustaka. Bagian ini mencakup penelitian-penelitian relevan sebelumnya serta teori-teori yang menjadi dasar untuk memahami analisis sentimen berbasis aspek
- BAB III : METODE PENELITIAN**  
 BAB III menguraikan metodologi penelitian yang diterapkan secara sistematis, dimulai dari proses akuisisi dataset hingga tahap akhir yang meliputi klasifikasi dan evaluasi model
- BAB IV : HASIL DAN ANALISIS**  
 BAB IV menyajikan hasil penelitian dan pembahasan. Paparan pada bab ini mencakup hasil akhir dari sistem yang telah dibangun, rincian hasil klasifikasi pada data uji, serta analisis performa sistem berdasarkan tingkat akurasinya.
- BAB V : KESIMPULAN DAN SARAN**  
 BAB V merupakan bab penutup yang terdiri dari kesimpulan dan saran. Kesimpulan menyajikan intisari dari hasil penelitian untuk menjawab tujuan yang telah dirumuskan, sementara saran

berisi rekomendasi untuk pengembangan penelitian selanjutnya.



## BAB II

### TINJUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Analisis sentimen atau *opinion mining* merupakan proses memahami dan mengekstrak data tekstual untuk memperoleh informasi mengenai sentimen yang terkandung di dalamnya, baik positif maupun negatif. Tujuan dari analisis ini adalah untuk mengetahui tingkat kepuasan publik serta pandangan konsumen terhadap suatu produk atau layanan, sehingga hasilnya dapat menjadi bahan evaluasi penting bagi penyedia layanan untuk meningkatkan kualitas.

Pendekatan klasik dalam analisis sentimen banyak memanfaatkan algoritma *machine learning* seperti *Support Vector Machine* (SVM). Pada tahapannya, teks ulasan terlebih dahulu melalui proses *pre-processing* sebelum diubah menjadi representasi numerik menggunakan metode pembobotan, misalnya *Term Frequency-Inverse Document Frequency* (TF-IDF). Penelitian analisis sentimen terhadap layanan Gojek di Twitter yang menerapkan metode SVM menunjukkan hasil klasifikasi sentimen positif dan negatif dengan akurasi keseluruhan mencapai 79,19% (Fitriyah dkk., 2020).

Perkembangan selanjutnya menunjukkan bahwa model *deep learning* berbasis arsitektur *Transformer*, seperti BERT (*Bidirectional Encoder Representations from Transformers*), memiliki performa lebih unggul dibanding metode klasik. BERT mampu menangkap hubungan kontekstual antar kata secara dua arah (*bidirectional*), sehingga pemahaman terhadap makna kalimat menjadi lebih mendalam. Keunggulan utama BERT terletak pada proses *pre-training* menggunakan data berskala besar, yang kemudian diikuti dengan *fine-tuning* pada dataset spesifik sesuai tugas, seperti klasifikasi sentimen. Penelitian yang menerapkan varian BERT, yakni SmallBERT, pada ulasan hotel berbahasa Indonesia berhasil mencapai akurasi sebesar 91,40% setelah melalui proses *fine-tuning* (Chandradev dkk., 2023).

Meskipun model seperti BERT telah terbukti efektif dalam klasifikasi sentimen umum, kebutuhan analisis yang lebih mendalam mendorong

pengembangan Analisis Sentimen Berbasis Aspek (*Aspect-Based Sentiment Analysis*). ABSA tidak hanya mengidentifikasi sentimen keseluruhan, tetapi juga mengevaluasi sentimen terhadap aspek-aspek spesifik dari suatu produk atau layanan. Seperti penelitian yang dilakukan pada aplikasi KAI *Access* menggunakan metode SVM untuk menganalisis aspek *Learnability*, *Efficiency*, *Errors*, dan *Satisfaction*, dengan hasil akurasi terbaik masing-masing sebesar 94,73%, 94,38%, 85,13%, dan 87,26% (Radiena dan Nugroho, 2023).

Untuk menangani teks berbahasa Indonesia secara lebih efektif, dikembangkan model *transformer* khusus seperti IndoBERT. IndoBERT merupakan model yang telah melalui *pre-training* dengan lebih dari 220 juta kata berbahasa Indonesia yang bersumber dari Wikipedia dan artikel berita. Pelatihan khusus ini memberikan IndoBERT pemahaman yang lebih mendalam terhadap struktur dan nuansa bahasa Indonesia. Dalam penelitian ABSA pada aplikasi M-Paspor, aspek diekstraksi menggunakan metode LDA (*Latent Dirichlet Allocation*), sedangkan klasifikasi sentimen dilakukan menggunakan IndoBERT. Hasil penelitian tersebut menunjukkan bahwa pemodelan aspek mencapai akurasi tertinggi sebesar 94% (Widiansyah dkk., 2024).

Analisis terhadap penelitian terdahulu menunjukkan bahwa penerapan ABSA pada layanan publik di Indonesia masih menghadapi keterbatasan, terutama karena pendekatan yang digunakan cenderung multi-tahap, dimana ekstraksi aspek dan klasifikasi sentimen dengan SVM atau IndoBERT dilakukan secara terpisah serta hanya terbatas pada domain tertentu. Celah penelitian ini membuka peluang untuk mengembangkan model end-to-end yang mampu mendeteksi aspek sekaligus menentukan sentimen dalam satu proses terpadu. Dalam konteks ini, IndoBERT dipilih karena arsitektur *transformernya* terbukti lebih unggul dibanding metode klasik seperti SVM, dan telah melalui *pre-training* pada korpus Bahasa Indonesia berskala besar, sehingga memiliki pemahaman yang mendalam terhadap struktur serta nuansa bahasa lokal. Keunggulan tersebut menjadikan IndoBERT lebih baik untuk dikembangkan menjadi model ABSA *end to end* yang efektif, efisien, dan relevan dalam menganalisis ulasan pengguna aplikasi layanan publik di Indonesia.

## 2.2 DASAR TEORI

### 2.2.1 Analisis Sentimen

Analisis sentimen adalah sebuah area dalam Pengolahan Bahasa Alami (NLP) yang bertujuan menelaah teks untuk menemukan pandangan atau emosi yang terkandung di dalamnya. Melalui proses ini teks seperti komentar pengguna dapat secara otomatis dikategorikan untuk menentukan polaritasnya, apakah cenderung positif, negatif, atau netral (Rahman dkk., 2024). Analisis sentimen telah diimplementasikan pada berbagai bidang, termasuk *e-commerce*, media sosial, dan layanan publik, di mana perusahaan atau penyedia layanan dapat memperoleh wawasan tentang kepuasan pengguna melalui ulasan yang mereka berikan (Nuriani dkk., 2024). Pendekatan yang digunakan dalam analisis sentimen umumnya dibagi menjadi tiga kategori utama, yaitu berbasis kamus (*lexicon-based*), pembelajaran mesin (*machine learning*), dan pembelajaran mendalam (Chinnasamy dkk., 2022).

### 2.2.2 Aspek Penelitian

Dalam konteks *Aspect Base Sentiment Analysis* (ABSA) aspek merupakan atribut, fitur, atau komponen spesifik dari suatu entitas baik itu produk, layanan, atau topik, yang menjadi target dari opini atau sentimen pengguna (Truşcă & Frasincar, 2023). Sebagai contoh pada ulasan sebuah restoran aspek dapat berupa layanan, makanan, dan tempat.

Untuk memastikan kategori aspek yang digunakan dalam penelitian ini bersifat objektif, terstruktur, dan relevan dengan domain rekayasa perangkat lunak, aspek yang didefinisikan tidak dilakukan secara manual dari data melainkan menggunakan dan mengadopsi karakteristik kualitas dari model ISO/IEC 25010 (*International Organization for Standardization/ International Electrotechnical Commission*) sebagai kerangka kerja formal untuk mendefinisikan aspek yang akan dianalisis. ISO/IEC 25010 merupakan bagian dari seri SQuaRE (*Systems and software Quality Requirements and Evaluation*) standar ini menyediakan sebuah model kualitas produk yang komprehensif untuk mengevaluasi kualitas perangkat lunak. Pendekatan ini sudah pernah dilakukan pada penelitian (Fehle dkk., 2025)

yang telah berhasil memetakan ulasan teks pengguna ke dalam atribut menggunakan ISO/IEC 25010.

### **2.2.3.1 Usability**

*Usability* didefinisikan sebagai sejauh mana suatu produk atau sistem dapat digunakan oleh pengguna tertentu untuk mencapai tujuan yang ditetapkan dengan efektivitas, efisiensi, dan kepuasan dalam konteks penggunaan tertentu (International Organization for Standardization, 2011). Dalam konteks penelitian ini aspek *Usability* mencakup semua ulasan pengguna yang membahas kemudahan dalam mempelajari, mengoperasikan, dan berinteraksi dengan antarmuka (UI/UX), serta kejelasan alur fungsionalitas aplikasi. Contoh ulasan yang menggambarkan aspek ini adalah “Aplikasinya mudah dipahami, tampilan simpel, jadi cepat ngerti cara pakainya” kalimat tersebut termasuk aspek *Usability*, karena membahas kemudahan memahami dan mengoperasikan aplikasi, serta memiliki sentimen positif karena pengguna merasa terbantu dan nyaman saat menggunakan aplikasi.

### **2.2.3.2 Reliability**

*Reliability* didefinisikan sebagai sejauh mana sistem, produk, atau komponen melakukan fungsi yang ditentukan dalam kondisi tertentu untuk periode waktu tertentu (International Organization for Standardization, 2011). Dalam penelitian ini aspek *Reliability* berfokus pada stabilitas sistem, mencakup sentimen pengguna terkait kegagalan fungsi (*crash*), *bug* program, kesalahan (*error*), waktu henti (*downtime*) server, dan ketidakmampuan fitur untuk berfungsi sebagaimana mestinya. Contoh ulasan yang sesuai adalah “Aplikasinya sering crash tiap buka halaman tertentu” kalimat tersebut memiliki aspek *reliability* karena menyoroti kegagalan aplikasi dalam berfungsi secara stabil, serta termasuk sentimen negatif karena pengguna mengalami gangguan berulang yang menyebabkan ketidaknyamanan.

### **2.2.3.3 Efficiency**

*Efficiency* didefinisikan sejauh mana kinerja yang relatif terhadap jumlah sumber daya yang digunakan dalam kondisi tertentu (International Organization for Standardization, 2011). Pada penelitian ini, aspek *Efficiency* akan mengidentifikasi

sentimen yang membahas kinerja aplikasi secara langsung, terutama terkait kecepatan, waktu respons, waktu muat (*loading time*), dan konsumsi sumber daya seperti boros baterai atau kuota yang dirasakan oleh pengguna. Contoh ulasan yang merepresentasikan aspek ini adalah “Aplikasinya lemot banget pas dibuka, loading-nya lama” kalimat tersebut termasuk aspek *Efficiency*, karena membahas kecepatan dan waktu respons sistem, serta menunjukkan sentimen negatif karena pengguna merasakan kinerja aplikasi yang lambat.

### 2.2.3 *Aspect-Based Sentiment Analysis (ABSA)*

Analisis sentimen berbasis aspek adalah sebuah tingkatan analisis yang lebih mendalam dari analisis sentiment konvensional, yang bertujuan untuk mengekstraksi dan mengidentifikasi opini atau sentimen terhadap aspek-aspek atau fitur-fitur spesifik dari suatu entitas, seperti produk atau layanan. Kebutuhan akan ABSA muncul karena ulasan pelanggan seringkali mengandung banyak aspek, di mana masing-masing aspek tersebut dapat memiliki polaritas sentimen yang berbeda (positif, negatif, atau netral). Sebagai contoh, sebuah ulasan restoran bisa menyatakan "Masakan dan lokasinya enak, tapi layanan yang diberikan tidak memuaskan" yang mengandung sentimen positif untuk aspek masakan dan tempat sekaligus sentimen negatif untuk aspek layanan. Dengan demikian, ABSA memungkinkan pemahaman yang lebih rinci mengenai kelebihan dan kekurangan suatu entitas di mata konsumen. (Amien dkk., 2021).

Proses ABSA secara tradisional melibatkan dua sub-tugas utama yang dieksekusi secara berurutan, meskipun pendekatan modern cenderung menggabungkannya.

#### 1. *Aspect Term Extraction*

Ekstraksi aspek merupakan komponen penting dalam *Sentiment Analysis* pada level aspek karena dapat menghasilkan analisis yang lebih akurat. Tidak hanya sekadar mengelompokkan kata pada setiap subtopik, proses ini bertujuan untuk mengidentifikasi aspek tertentu dari suatu produk atau layanan yang memperoleh tanggapan positif maupun negatif dari pengguna. Secara umum, ekstraksi aspek berfokus pada pemetaan frasa dan kata benda. Beberapa

pendekatan yang dapat digunakan dalam ekstraksi aspek antara lain *supervised learning*, *unsupervised learning*, dan *rule-based learning*. Pendekatan *supervised learning* memerlukan data berlabel, sedangkan *unsupervised learning* dapat dilakukan tanpa pelabelan data (Adimanggala dkk., 2021).

## 2. *Sentiment Polarity Classification*

Setelah aspek berhasil diekstraksi, tahap berikutnya adalah menentukan polaritas sentimen yang terkait dengan aspek tersebut. Sentimen kemudian dikategorikan ke dalam kelas positif, negatif, atau netral. Proses ini dapat dilakukan menggunakan berbagai algoritma *machine learning* maupun pendekatan berbasis *deep learning*, seperti *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *k-Nearest Neighbor* (k-NN), serta model berbasis *transformer* yang lebih modern. Data berlabel biasanya diperlukan untuk melatih model, sementara performa klasifikasi dievaluasi menggunakan metrik standar, seperti *accuracy*, *precision*, *recall*, dan *F-measure* (Iskandar dan Nataliani, 2021).

### 2.2.4 Aplikasi Pelayanan Masyarakat

Aplikasi pelayanan masyarakat merupakan perwujudan dari *e-government*, yaitu pemanfaatan teknologi informasi oleh pemerintah untuk menyelenggarakan pemerintahan dan menyediakan layanan publik yang lebih efektif, efisien, dan transparan (Muliawaty & Hendryawan, 2020). Beberapa aplikasi yang termasuk di dalamnya adalah Mobile JKN (Jaminan Kesehatan Nasional), Samsat Digital (SIGNAL), dan M-Paspor. Mobile JKN merupakan aplikasi resmi dari BPJS Kesehatan yang memudahkan peserta Jaminan Kesehatan Nasional dalam mengakses berbagai layanan kesehatan secara digital. Samsat Digital Nasional (SIGNAL) adalah aplikasi yang dikembangkan oleh Korps Lalu Lintas Kepolisian Negara Republik Indonesia (Korlantas Polri) untuk memudahkan masyarakat membayar pajak kendaraan bermotor secara aman dan mudah. Sementara itu, M-Paspor adalah aplikasi imigrasi yang dikembangkan Direktorat Jenderal Imigrasi untuk mempermudah masyarakat dalam melakukan permohonan dan pengelolaan paspor.

Analisis sentimen atau *opinion mining* menjadi sangat penting untuk memahami opini pengguna dari ulasan yang mereka berikan melalui platform seperti Google Play Store. Ulasan pengguna mengandung informasi berharga mengenai pengalaman nyata mereka, baik dalam bentuk kepuasan maupun keluhan (Filemon Haganta Kaban dan Yudistira, 2021). Analisis sentimen dapat memberikan wawasan yang lebih mendalam dibandingkan analisis deskriptif konvensional, membantu mengidentifikasi masalah teknis secara dinamis, serta menjadi alat evaluasi bagi penyelenggara layanan. Hasil analisis bahkan dapat dimanfaatkan sebagai sistem peringatan dini agar keluhan bernuansa negatif, misalnya terkait gangguan server, kendala verifikasi, atau masalah kecepatan layanan, dapat segera diprioritaskan penanganannya. Penelitian ini berfokus pada penerapan Aspect-Based Sentiment Analysis (ABSA) berbasis IndoBERT terhadap ulasan pengguna Mobile JKN, SIGNAL, dan M-Paspor. Dengan ABSA, analisis tidak hanya berhenti pada identifikasi polaritas sentimen, tetapi juga menghubungkannya dengan aspek spesifik seperti *usability*, *reliability*, dan *efficiency*, sehingga dapat memberikan gambaran yang lebih komprehensif mengenai kualitas layanan digital pemerintah.

### **2.2.5 Natural Language Processing (NLP)**

*Natural Language Processing* (NLP) adalah cabang ilmu komputer dan kecerdasan buatan yang mempelajari bagaimana komputer dapat memahami, menganalisis, dan memproses bahasa manusia secara otomatis dan efisien. Dalam penerapannya pada Bahasa Indonesia, NLP meliputi berbagai teknik dasar seperti *stemming*, *part-of-speech tagging*, serta analisis sentimen. Penerapan NLP membantu menyederhanakan teks yang kompleks dan meningkatkan efektivitas pengolahan data berbasis bahasa. Proses kerja NLP biasanya diawali dengan tahap *preprocessing* untuk membersihkan teks, kemudian dilanjutkan dengan *tokenisasi* guna memecah teks menjadi unit-unit kecil, serta penggunaan model klasifikasi atau metode ekstraksi informasi. Dengan adanya perkembangan teknologi seperti pembelajaran mesin dan jaringan saraf, kemampuan NLP dalam memahami konteks kalimat semakin baik dibandingkan dengan metode statistik tradisional.

Saat ini, NLP banyak dimanfaatkan di berbagai bidang salah satunya dalam analisis sentimen yang digunakan untuk mengetahui opini publik di media sosial, maupun pada aplikasi lain yang memerlukan pengolahan bahasa alami secara otomatis (M. Amien, 2023).

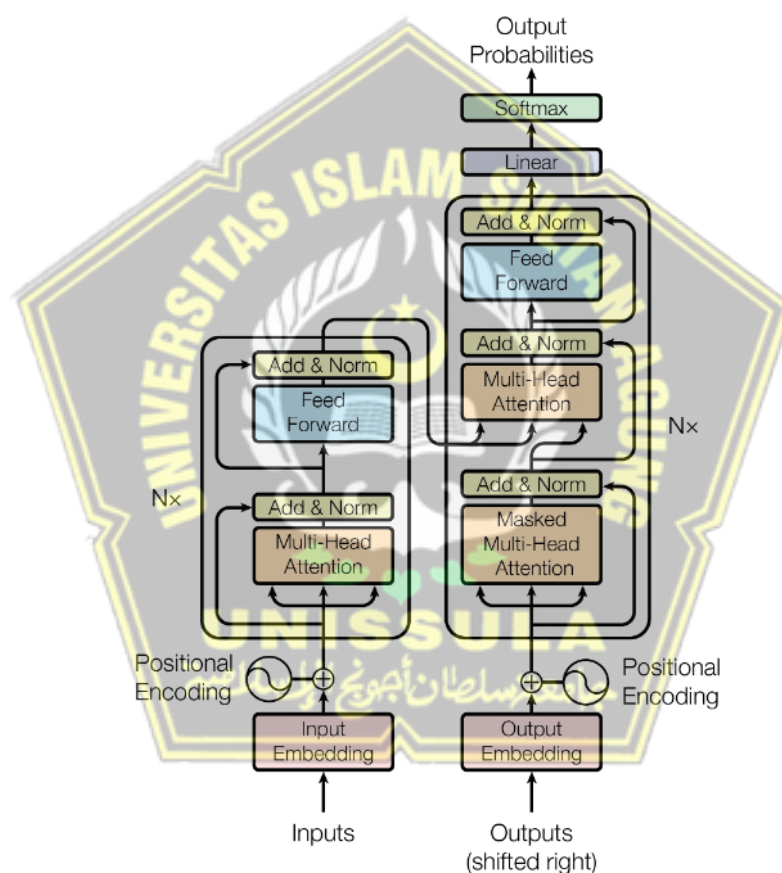
### 2.2.6 Deep Learning

*Deep learning* merupakan salah satu cabang dari *machine learning* yang memanfaatkan arsitektur *neural network* berlapis untuk mempelajari representasi data secara hierarkis. Tidak seperti *machine learning* tradisional yang sangat bergantung pada proses ekstraksi fitur secara manual, *deep learning* mampu melakukan *automatic feature extraction*, sehingga dapat menghasilkan performa dengan akurasi yang lebih baik pada data berukuran besar dan kompleks (Zhou dkk., 2019). Dalam konteks pengolahan bahasa alami (*Natural Language Processing*), *deep learning* mengubah data teks menjadi representasi numerik dalam bentuk vektor atau matriks yang dapat diproses oleh algoritma. Arsitektur awal seperti *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM) banyak digunakan untuk memproses data sekuensial karena kemampuannya menangkap dependensi antar kata dalam sebuah kalimat (Naquitasia dkk., 2022). Namun, RNN dan LSTM memiliki keterbatasan dalam memahami hubungan kontekstual jangka panjang antar kata yang posisinya berjauhan.

Arsitektur *Transformer* yang didasarkan pada mekanisme *self-attention*. *Transformer* memungkinkan model memahami konteks global dalam sebuah kalimat secara paralel, sehingga lebih efisien dan efektif dibandingkan arsitektur sekuensial. Perkembangan ini melahirkan berbagai model berbasis *pre-trained language model* seperti BERT dan turunannya, termasuk IndoBERT yang secara khusus dikembangkan untuk bahasa Indonesia. Dengan kemampuan representasi kontekstual yang lebih baik, *deep learning* berbasis *Transformer* menjadi pendekatan utama dalam berbagai tugas NLP modern, termasuk analisis sentimen berbasis aspek (ABSA).

### 2.2.7 Transformers

*Transformers* dirancang untuk mengatasi keterbatasan model sebelumnya seperti RNN dan LSTM, terutama dalam menangani urutan data yang panjang. Inovasi utama dari model ini adalah penggunaan mekanisme *self-attention*, yang memungkinkan model mempertimbangkan hubungan antar kata dalam kalimat secara bersamaan, sehingga meningkatkan pemahaman konteks. Gambar 2.1 berikut menggambarkan arsitektur *Transformers* (Vaswani dkk., 2017).



Gambar 2. 1 Arsitektur model *transformers* (Vaswani dkk., 2017)

Arsitektur *Transformers* terdiri atas dua komponen utama, yaitu *encoder* dan *decoder* dengan lapisan layer yang memiliki bloks *self-attention* dan lapisan *feed forward*.

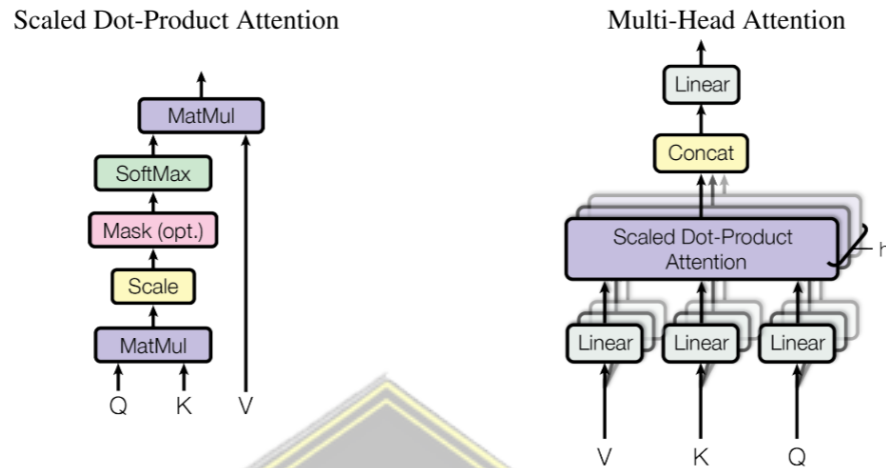
#### a. *Encoder dan Decoder*

*Encoder* berperan dalam memproses keseluruhan teks masukan secara paralel. Arsitektur *encoder* dibangun dari sejumlah lapisan identik yang ditumpuk

secara berulang. Setiap lapisan mencakup dua sub-lapisan utama, yaitu *self-attention mechanism* yang berfungsi menangkap hubungan antar-token dalam teks, serta *feed-forward neural network* yang bertugas melakukan transformasi non-linear terhadap representasi yang dihasilkan (Han dkk., 2021). Sedangkan decoder berfungsi untuk mengubah representasi yang dihasilkan *encoder* menjadi urutan keluaran prediktif. Setiap lapisannya terdiri atas tiga komponen, yaitu *self-attention*, *cross-attention*, dan *feed-forward neural network*. *Self-attention* memungkinkan model memperhatikan token lain pada keluaran parsial, sedangkan *cross-attention* menghubungkan informasi dari *encoder* agar konteks input tetap terjaga (Mohiuddin dkk., 2023). Dalam analisis sentimen berbasis aspek (ABSA), mekanisme ini membantu model membedakan makna kata berdasarkan aspek. Misalnya, kata “cepat” dapat merujuk pada kecepatan pelayanan (aspek *efficiency*) atau kecepatan respon keluhan (aspek *reliability*), sehingga decoder mampu menghasilkan representasi yang lebih kontekstual.

b. *Scaled Dot-Product Attention*

Input pada mekanisme perhatian terdiri dari *query*, *key*, dan *value*. Prosesnya dilakukan dengan mencocokkan *query* terhadap seluruh *key*, lalu hasilnya dinormalisasi menggunakan fungsi *softmax* untuk menghasilkan bobot perhatian pada setiap *value*. Secara umum, terdapat dua jenis perhatian yang banyak digunakan, yaitu *additive attention* dan *dot-product attention*. Mekanisme *dot-product attention* yang dipakai dalam *Transformer* sebenarnya serupa dengan metode sebelumnya, hanya berbeda pada cara penskalaannya. Alur kerja dari mekanisme *Multi-Head Attention* dalam *Transformer* ditunjukkan pada Gambar 2.2.



Gambar 2. 2 *Scaled Dot-Product Attention dan Multi-head attention (Mohiuddin dkk., 2023)*

c. *Multi-head attention layer*

Dalam *Vision Transformer* (ViT), terdapat lapisan yang berfungsi untuk memproses informasi dari citra, yaitu *multi-head attention*. Mekanisme ini terdiri atas beberapa unit *attention* yang dijalankan secara paralel, sehingga model mampu menangkap informasi dari berbagai representasi maupun posisi secara bersamaan. Secara sederhana, *multi-head attention* memungkinkan setiap *patch* gambar saling berinteraksi dengan cara bertukar informasi melalui beberapa jalur secara paralel. Proses ini memanfaatkan *scaled dot-product attention* dengan komponen *query*, *key*, dan *value* untuk menghasilkan representasi yang lebih kaya dan kontekstual (Vaswani dkk., 2017).

d. *Position-wise Feed-Forward Networks*

Selain sublapisan perhatian, setiap lapisan pada *encoder* maupun *decoder* juga dilengkapi dengan jaringan *feed-forward* yang terhubung penuh. Jaringan ini diterapkan pada setiap posisi secara terpisah namun identik. Struktur *feed-forward* tersebut terdiri atas dua transformasi linier yang dipisahkan oleh fungsi aktivasi ReLU. Meskipun transformasi liniernya konsisten di semua posisi, parameter yang digunakan berbeda pada tiap lapisan. Secara alternatif, mekanisme ini dapat dipandang sebagai dua operasi konvolusi dengan kernel satu dimensi. Dalam arsitektur *Transformer*, dimensi masukan dan keluaran

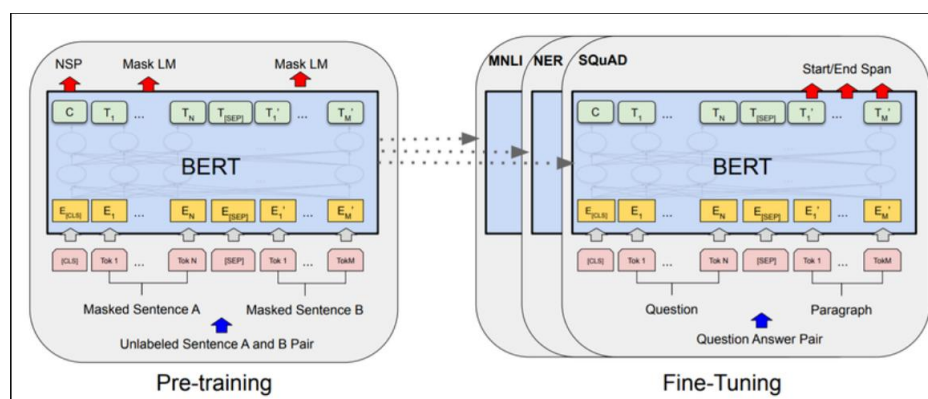
ditetapkan sebesar 512, sedangkan dimensi lapisan tersembunyi mencapai 2048 (Vaswani dkk., 2017).

e. *Positional Encoding*

*Positional Encoding* atau *Positional Embedding* merupakan mekanisme yang digunakan dalam model *Transformer*, termasuk IndoBERT, untuk memberikan informasi mengenai urutan kata dalam sebuah kalimat. Mekanisme ini penting karena arsitektur *self-attention* pada *Transformer* tidak secara inheren memahami posisi kata, berbeda dengan model berurutan seperti RNN yang memproses kata satu per satu. Dengan adanya *positional embedding*, IndoBERT dapat mempertahankan informasi urutan sehingga representasi kata yang dihasilkan lebih sesuai dengan struktur kalimat dalam bahasa Indonesia (Vaswani dkk., 2017).

## 2.2.8 BERT

*Bidirectional Encoder Representations from Transformers* (BERT) adalah sebuah model untuk pemrosesan bahasa alami yang diperkenalkan oleh Google AI pada tahun 2018. Model ini telah mengubah paradigma dalam pemrosesan bahasa alami (NLP) dengan pendekatan inovatif dalam memahami konteks teks. Berbeda dengan model-model sebelumnya, BERT memanfaatkan arsitektur *Transformer* dan menerapkan pelatihan *bidirectional*, yang memungkinkan model untuk memahami hubungan antar kata dalam kalimat berdasarkan konteksnya (Devlin dkk., 2019). BERT dua tahap utama yaitu *pre-training* dan *fine tuning*.



Gambar 2. 3 *Pre-training dan fine-tuning* (Devlin dkk., 2019)

Gambar 2.3 mengilustrasikan proses pre-training dan fine-tuning. Berikut adalah penjelasan untuk setiap tahapnya:

### 1. *Pre-training*

*Pre-training* merupakan pendekatan awal yang digunakan untuk melatih BERT agar mampu membaca dan memahami teks beserta konteksnya secara lebih mendalam. Secara umum arsitektur *pre-training* dan *fine-tuning* pada BERT perbedaan utamanya terletak pada lapisan *output*. Setiap input yang dimasukkan ke dalam model dipecah terlebih dahulu menjadi token-token yang dapat berupa kata, frasa, maupun karakter. Seperti yang ditunjukkan pada Gambar 2.3, token khusus [CLS] selalu ditempatkan di awal input, sementara token [SEP] ditambahkan di akhir untuk memisahkan dua segmen teks, misalnya pada pasangan pertanyaan dan jawaban (Tandijaya dkk., 2021). Pada tahap pre-training BERT menggunakan dua tujuan pelatihan *unsupervised* sekaligus, yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP).

#### a. *Masked Language Modeling* (MLM)

Berbeda dengan *transformer* yang memanfaatkan mekanisme *self-attention* untuk memproses seluruh urutan token sekaligus dalam memahami konteks dari kedua arah (kiri maupun kanan), BERT menambahkan langkah khusus dengan menutup (*masking*) sebagian token input secara acak dan meminta model memprediksi token yang ditutup tersebut berdasarkan konteks di sekitarnya. Proses ini dikenal sebagai *Masked Language Modeling* (MLM). Dalam praktiknya, sekitar 15% dari token pada input diganti dengan token khusus [MASK]. Selanjutnya, BERT dilatih untuk menebak kata asli yang tersembunyi tersebut dengan mempertimbangkan informasi dari token lain yang tidak ditutup dalam urutan kata. Dengan cara ini, MLM memungkinkan BERT untuk memperoleh representasi yang benar-benar bidirectional sekaligus memahami keterkaitan antar kata dalam sebuah kalimat secara lebih mendalam (Devlin dkk., 2019).

#### b. *Next Sentence Prediction* (NSP)

Untuk menyelesaikan tugas-tugas tertentu, BERT dilatih selama fase *pre-training* dengan proses yang disebut *Next Sentence Prediction* (NSP).

Tujuannya adalah agar model dapat memahami hubungan dan konteks antar kalimat. Melalui NSP, BERT diajarkan untuk memprediksi apakah sebuah kalimat yang merupakan kelanjutan dari kalimat sebelumnya. Proses ini menghasilkan dua kemungkinan hasil dengan probabilitas 50%, yaitu label 'IsNext' jika kalimat kedua adalah lanjutan dari kalimat pertama, atau 'NotNext' jika bukan (Devlin dkk., 2019).

## 2. *Fine-tuning*

Pada tahap ini *fine-tuning*, BERT disesuaikan untuk berbagai tugas pemrosesan bahasa alami (NLP) menggunakan dataset berlabel yang lebih kecil, seperti klasifikasi teks (*Multi-Genre Natural Language Inference/MNLI*), pengenalan entitas dalam teks (*Named Entity Recognition/NER*), dan pencarian jawaban dalam teks (*Stanford Question Answering Dataset/SQuAD*). Salah satu Kelebihan paling signifikan dari BERT adalah pemahaman kontekstualnya yang mendalam. Secara *bidirectional*, yang membuatnya lebih efektif dalam menangkap makna kata dibandingkan dengan model-model NLP sebelumnya (Devlin dkk., 2019).

Dengan demikian, kemampuan BERT dalam memahami konteks secara *bidirectional* dan fleksibilitasnya dalam dioptimalkan untuk berbagai tugas spesifik menjadikannya sebagai model yang sangat efektif dalam berbagai aplikasi pemrosesan bahasa alami (NLP), termasuk analisis sentimen, klasifikasi teks, pengenalan entitas, dan pencarian jawaban.

### 2.2.9 IndoBert

IndoBERT (*Indonesian Bidirectional Encoder Representations from Transformers*) merupakan model berbasis arsitektur *Transformer* bergaya BERT yang dilatih sepenuhnya dengan pendekatan *masked language modeling* menggunakan *framework Huggingface*, dengan konfigurasi standar BERT-Base (*uncased*). Model ini terdiri atas 12 lapisan tersembunyi dengan dimensi 768, 12 *attention heads*, serta lapisan *feed-forward* berukuran 3.072. *Framework Huggingface* dimodifikasi agar mampu membaca berbagai aliran teks dari beragam blok dokumen, dengan panjang 512 token per *batch* saat pelatihan. Kosakata yang

digunakan berjumlah 31.923 *WordPiece* khusus bahasa Indonesia. Data latih IndoBERT bersumber dari tiga korpus utama: Web Corpus Indonesia (90 juta kata), artikel berita dari Kompas, Tempo, dan Liputan6 (55 juta kata), serta Wikipedia Bahasa Indonesia (74 juta kata). Setelah melalui tahap pra-proses menjadi blok teks bertoken sepanjang 512, diperoleh 1.067.581 contoh latih dan 13.985 contoh validasi (tanpa duplikasi). Pelatihan dilakukan dengan empat GPU Nvidia V100 berkapasitas 16GB, ukuran *batch* 128, laju pembelajaran  $1e-4$ , optimisasi Adam, dan *linear scheduler*. Model dilatih selama dua bulan, mencakup 2,4 juta langkah (180 epoch). Hasilnya, model mencapai nilai *perplexity* akhir sebesar 3,97 pada set pengembangan (Koto dkk., 2020).

Dalam proses pengembangan IndoBERT, peneliti menggunakan teknik *pre-training* yang serupa dengan BERT yaitu *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Melalui MLM, model dilatih untuk menebak kata yang hilang dalam sebuah kalimat dengan mempertimbangkan konteks di sekitarnya, sedangkan NSP berfungsi untuk melatih model dalam memahami keterkaitan antar kalimat. Temuan penelitian menunjukkan bahwa IndoBERT secara konsisten melampaui performa *Multilingual BERT (M-BERT)* pada berbagai tugas NLP yang berbasis bahasa Indonesia. Keunggulan utama IndoBERT terletak pada kemampuannya menangkap nuansa dan konteks bahasa Indonesia dengan lebih baik, sehingga dapat diimplementasikan dalam beragam aplikasi, mulai dari klasifikasi teks, analisis sentimen, ekstraksi informasi, penerjemahan, hingga pembuatan teks. Dengan keunggulan tersebut, IndoBERT diharapkan mampu mendukung pengolahan data berbahasa Indonesia secara lebih akurat dan efisien (Hidayat dan Pramudita, 2024).

#### **2.2.10 Confusion Matrix**

*Confusion matrix* merupakan teknik evaluasi kinerja model klasifikasi dalam *machine learning* yang menampilkan informasi mengenai perbandingan antara hasil klasifikasi sebenarnya dengan hasil prediksi yang dihasilkan oleh sistem (Hakim dkk., 2024).

Perbandingan antara kelas prediksi yang dihasilkan algoritma dengan data aktual selanjutnya dimanfaatkan untuk menghitung metrik evaluasi, meliputi *accuracy*, *precision*, *recall*, dan *f1-score*, guna menilai performa model klasifikasi.

1. *Accuracy* merupakan proses pengukuran pada seberapa banyak prediksi yang benar dibandingkan dengan seluruh data.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (1)$$

2. *Precision* merupakan proses mengukur sejauh mana prediksi positif yang dihasilkan pada model benar

$$Precision = \frac{Tp}{Tp + Fp} \quad (2)$$

3. *Recall* merupakan proses mengukur sejauh mana model dapat menangkap semua data positif yang sebenarnya.

$$Recall = \frac{Tp}{Tp + Fn} \quad (3)$$

4. F1-score didefinisikan sebagai rata-rata harmonik dari *precision* dan *recall*. yang bertujuan untuk menyeimbangkan antara *precision* dan *recall*.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan:

*True Positive* (TP) : Klasifikasi dengan hasil positif pada data asli yang positif

*True Negative* (TN) : Klasifikasi hasil negatif pada data asli yang negatif

*False Positive* (FP) : Klasifikasi dengan hasil positif pada data asli yang negatif

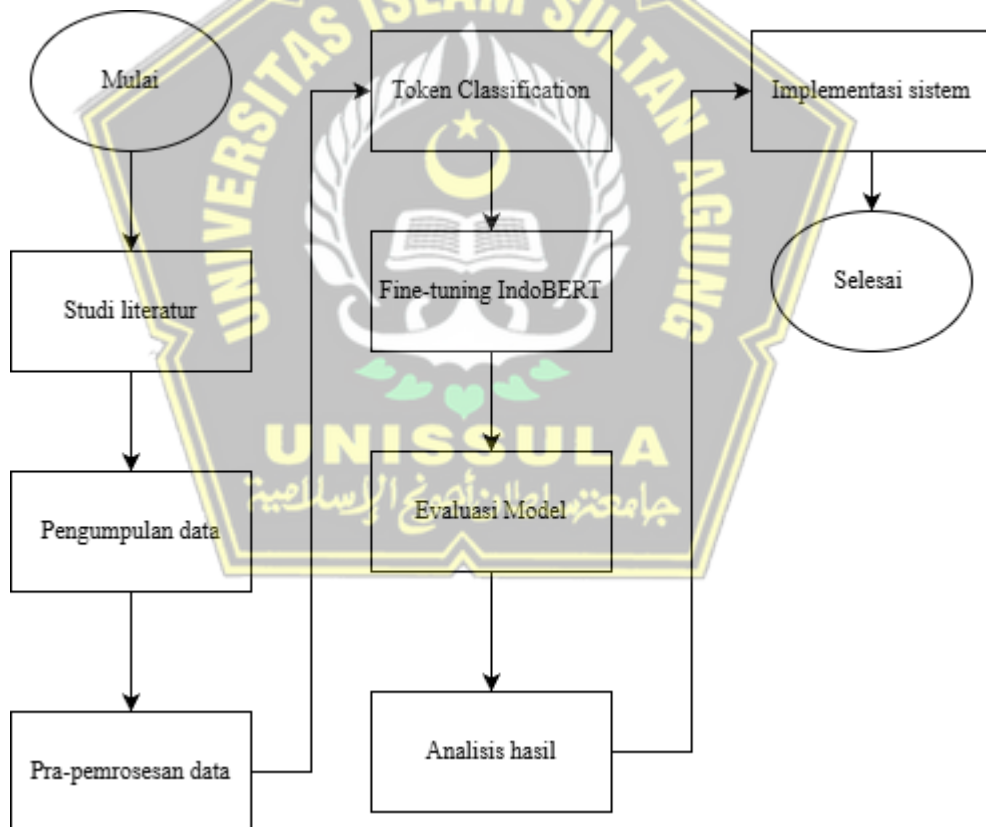
*False Negative* (FN) : Klasifikasi dengan hasil negatif pada data asli yang positif

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Metodologi Penelitian

Penelitian ini bertujuan untuk merancang dan membangun sistem analisis sentimen berbasis aspek pada ulasan aplikasi pelayanan masyarakat dengan menggunakan model IndoBERT. Untuk mencapai tujuan tersebut, penelitian ini dimulai dari studi literatur untuk membangun landasan teori, diikuti dengan pengumpulan dan persiapan data, perancangan dan pelatihan model, hingga implementasi model ke dalam sebuah sistem. Alur metodologi dapat dilihat pada gambar berikut.



Gambar 3. 1 Alur metodologi penelitian

Seperti terlihat pada Gambar 3.1 tahapan penelitian dimulai dengan studi literatur untuk membangun landasan teori yang kuat. Tahap selanjutnya Adalah pengumpulan data ulasan aplikasi dari *Google Play Store*. Data mentah tersebut

kemudian melalui proses pra-pemrosesan data, pada tahap ini melakukan dua proses yaitu *preprocessing* data dan Pelabelan Data (*Labeling*) secara manual untuk menyiapkan dataset yang bersih dan terannotasi. Dataset ini kemudian digunakan dalam tahap *Token Classification*, untuk memformulasikan *Aspect-Based Sentiment Analysis* (ABSA) pada level token. BIO (*Begin, Inside, Outside*), sehingga memungkinkan pendeteksian lebih dari satu aspek dalam satu kalimat. Setelah itu dilakukan proses *Fine-tuning* dengan *hyperparameter* yang telah ditentukan. Performa model yang telah dilatih kemudian diukur secara kuantitatif pada tahap Evaluasi Model menggunakan *dataset training*, *dataset validation*, dan *dataset testing*. Setelah itu dilakukan analisis berdasarkan hasil evaluasi yang telah didapatkan. Pada tahap akhir model terbaik diimplementasikan ke dalam sebuah sistem berbasis web menggunakan *framework Streamlit*.

### 3.2 Studi Literatur

Tahap awal penelitian adalah melakukan studi literatur. Pada tahap ini, dilakukan pengumpulan, penelaahan, dan tinjauan terhadap sumber-sumber ilmiah seperti buku, artikel jurnal yang relevan dengan topik *Aspect-Based Sentiment Analysis* (ABSA). Tinjauan difokuskan pada penelitian terdahulu yang menerapkan model berbasis *Transformer*, khususnya BERT, untuk kasus teks berbahasa Indonesia. Hasil dari studi literatur menunjukkan bahwa model *pre-trained* seperti IndoBERT memiliki performa yang unggul untuk berbagai tugas *Natural Language Processing* (NLP) dalam Bahasa Indonesia karena telah dilatih pada korpus data yang besar dan beragam.

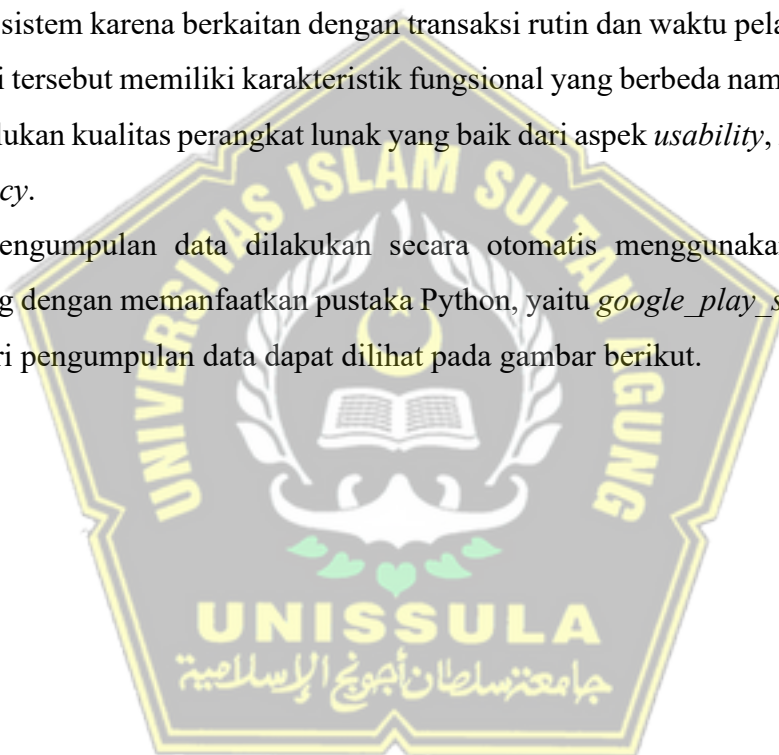
### 3.3 Pengumpulan Data

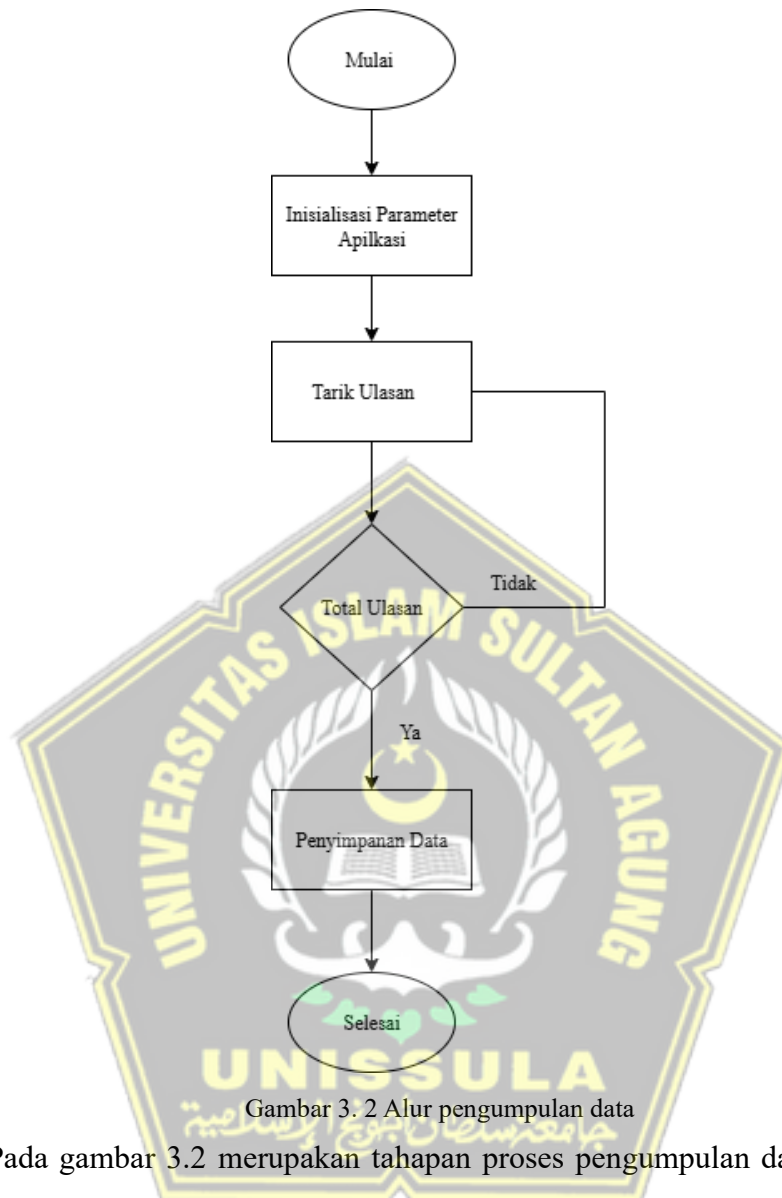
Data merupakan komponen krusial dalam penelitian berbasis *machine learning*. Data yang digunakan dalam penelitian ini adalah data ulasan pengguna yang bersumber dari aplikasi pelayanan masyarakat yang tersedia di Google Play Store. Untuk aplikasi yang dipilih yaitu Mobile JKN, M-Paspor, dan Samsat Digital Nasional (SIGNAL Pemilihan ketiga aplikasi ini tidak dilakukan secara acak, melainkan didasarkan pada pertimbangan representatif terhadap beragam sektor

layanan publik yang bersifat kritikal dan berdampak langsung pada masyarakat, yaitu layanan kesehatan, keimigrasian, dan administrasi kendaraan bermotor.

Mobile JKN merepresentasikan layanan publik di bidang kesehatan, yang memiliki tingkat ketergantungan tinggi terhadap keandalan sistem dan kontinuitas layanan. M-Paspor merepresentasikan layanan keimigrasian, yang menuntut ketepatan proses, kejelasan alur layanan, serta stabilitas sistem dalam mendukung kebutuhan administratif masyarakat. Sementara itu SIGNAL merepresentasikan layanan administrasi kendaraan bermotor, yang menekankan efisiensi proses dan kinerja sistem karena berkaitan dengan transaksi rutin dan waktu pelayanan. Ketiga aplikasi tersebut memiliki karakteristik fungsional yang berbeda namun sama-sama memerlukan kualitas perangkat lunak yang baik dari aspek *usability*, *reliability*, dan *efficiency*.

Pengumpulan data dilakukan secara otomatis menggunakan teknik web scraping dengan memanfaatkan pustaka Python, yaitu *google\_play\_scraper*. Untuk alur dari pengumpulan data dapat dilihat pada gambar berikut.





Gambar 3. 2 Alur pengumpulan data

Pada gambar 3.2 merupakan tahapan proses pengumpulan data ulasan dari aplikasi layanan publik menggunakan pustaka `google_play_scraper`. Berikut penjelasan dari setiap tahapan tersebut:

#### 1. Inisialisasi parameter

Tahap pertama adalah menentukan parameter yang diperlukan untuk melakukan proses pengambilan data, *parameter* yang digunakan meliputi:

Tabel 3. 1 *Parameter* pengambilan data

<i>Parameter</i>	<i>Value</i>	<i>Description</i>
app_id	'app.bpjs.mobile'	ID unik dari aplikasi di
	'app.signal.id'	Google Play Store.

<i>Parameter</i>	<i>Value</i>	<i>Description</i>
	'id.go.imigrasi.paspor_online'	Digunakan untuk menentukan sumber data ulasan. Masing-masing mewakili aplikasi Mobile JKN, M-Paspor, dan SIGNAL.
lang	id	Menentukan bahasa ulasan yang diambil, dalam hal ini Bahasa Indonesia, agar data relevan dengan konteks pengguna lokal.
country	id	Menentukan negara asal ulasan. Pengaturan 'id' memastikan hanya ulasan dari pengguna Indonesia yang dikumpulkan.
sort	Sort.NEWEST	Mengatur urutan pengambilan ulasan berdasarkan yang terbaru terlebih dahulu agar data mencerminkan opini terkini pengguna.
Count	200	Menentukan jumlah maksimum ulasan yang diambil dalam setiap batch proses scraping untuk efisiensi dan stabilitas sistem.

<i>Parameter</i>	<i>Value</i>	<i>Description</i>
<i>Num_reviews</i>	1700	Jumlah total target ulasan yang akan dikumpulkan dari masing-masing aplikasi.

## 2. Tarik ulasan

Setelah parameter ditentukan, sistem mulai melakukan proses pengambilan data menggunakan pustaka `google_play_scraper`. Pustaka ini berfungsi untuk mengekstrak ulasan pengguna dari *Google Play Store* secara otomatis. Proses dilakukan secara *batch* artinya data ditarik secara bertahap untuk efisiensi dan menghindari *request limit*.

## 3. Total ulasan

Pada tahap ini sistem memeriksa apakah jumlah ulasan yang sudah diambil telah mencapai target jumlah maksimum sesuai parameter awal. Jika belum tercapai sistem akan kembali ke tahap tarik ulasan untuk melanjutkan proses pengambilan *batch* berikutnya dan jika sudah terpenuhi proses dilanjutkan ke tahap penyimpanan data. Proses ini berlangsung secara iteratif hingga semua ulasan yang diinginkan berhasil dikumpulkan.

## 4. Penyimpanan Data

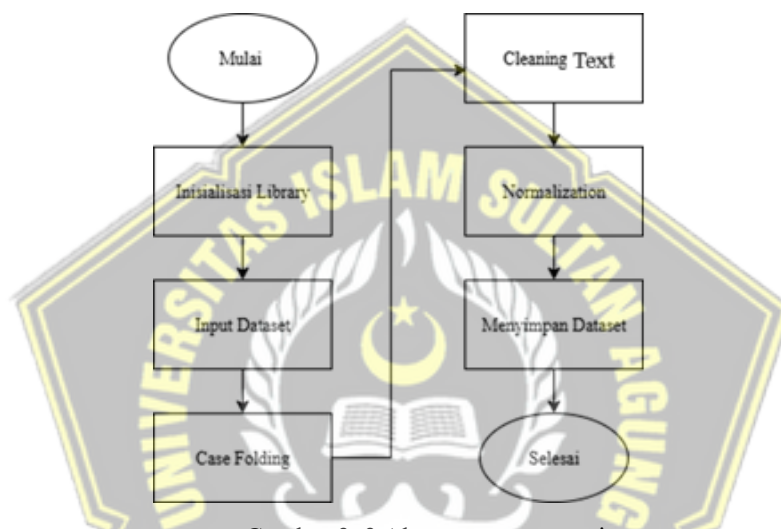
Setelah seluruh ulasan berhasil dikumpulkan, data yang ada di memori disimpan dalam format CSV agar mudah digunakan pada tahap berikutnya seperti *text preprocessing* dan pelatihan model. Isi teks ulasan format CSV dipilih karena fleksibel dan dapat langsung diproses menggunakan pustaka seperti `pandas` pada tahap analisis selanjutnya.

### 3.4 Pra-Pemrosesan Data

Pada tahap pra-pemrosesan data dilakukan dua langkah utama, yaitu *text processing* dan *labeling data*.

### 3.4.1 Text preprocessing

Pada *text preprocessing* melakukan persiapan data dari data mentah hingga siap digunakan, dilakukan beberapa tahapan yaitu *case folding* atau merubah teks menjadi huruf kecil untuk memastikan konsistensi data, melakukan *cleaning data* untuk menghapus kata yang tidak dibutuhkan seperti simbol, tanda baca, hastag, url, emoticon, angka, dan kata yang kurang dari dua huruf, dan melakukan *normalization* pada text untuk mengurangi variasi kata seperti “ga” menjadi “tidak”. Berikut merupakan alur dari *text preprocessing*.



Gambar 3. 3 Alur text preprocessing

Pada gambar 3.4 merupakan proses alur dari *text preprocessing* berikut penjelasan dari setiap alur tersebut:

1. Inisialisasi *library*

pada tahap inisialisasi pustaka (*library*) dalam proses *text preprocessing* yang digunakan antara lain:

  - a. Pandas digunakan untuk membaca, mengelola, dan menyimpan dataset dalam format tabel dalam hal ini format CSV.
  - b. *re* (*regular expression*) digunakan untuk proses *cleaning* dengan menghapus elemen yang tidak relevan seperti angka, simbol, dan tanda baca.
  - c. kamus normalisasi (*dictionary custom*) digunakan untuk mengganti kata tidak baku atau slang menjadi kata baku. Inisialisasi ini bertujuan agar seluruh *library* siap digunakan dalam tahap pemrosesan teks selanjutnya.

## 2. Input dataset

Pada tahap ini dilakukan proses penginputan dataset ulasan mentah yang diperoleh dari Google Play Store dalam format CSV. Dataset ini berisi teks ulasan pengguna yang masih bercampur dengan berbagai karakter seperti huruf besar, simbol, emoticon, dan bahasa tidak baku. Proses ini dilakukan menggunakan *library pandas* yang berfungsi untuk mengelola data tabular agar lebih mudah diolah.

## 3. Case folding

Tahapan ini bertujuan untuk menyeragamkan seluruh huruf dalam teks menjadi huruf kecil agar konsistensi data tetap terjaga. Proses ini penting agar kata yang sama dengan perbedaan kapitalisasi tidak dianggap berbeda oleh sistem. Implementasi dilakukan dengan fungsi *string* bawaan Python (`.lower()`) yang diterapkan pada setiap teks dalam kolom dataset. berikut merupakan hasil sebelum dan setelah *case folding*.

Tabel 3. 2 *Case folding*

<i>Before</i>	<i>After</i>
Aplikasi JKN ini bgs banget untuk cek antrean! 🙄	aplikasi jkn ini bgs banget untuk cek antrean! 🙄

## 4. Cleaning text

Tahap ini merupakan proses pembersihan teks dari elemen-elemen yang tidak relevan untuk analisis. Komponen yang dibersihkan meliputi simbol, tanda baca, URL, angka, hashtag, emoticon, serta kata yang panjangnya kurang dari dua huruf proses ini menggunakan *library re (regular expression)*. Dengan demikian, teks yang dihasilkan menjadi lebih bersih dan mudah diproses. berikut merupakan hasil sebelum dan setelah *cleaning text*.

Tabel 3. 3 *Cleaning text*

<i>Before</i>	<i>After</i>

aplikasi jkn ini bgs banget untuk cek antrean! 📌	aplikasi jkn ini bgs banget untuk cek antrean
--	---

#### 5. *Normalization*

Setelah teks dibersihkan, dilakukan proses *normalization* untuk mengganti kata tidak baku, singkatan, atau slang menjadi kata baku menggunakan kamus normalisasi. Proses ini memastikan setiap kata memiliki bentuk yang seragam sehingga memudahkan analisis selanjutnya. Berikut merupakan hasil sebelum dan setelah *normalization*.

Tabel 3. 4 *Normalization*

<i>Before</i>	<i>After</i>
aplikasi jkn ini bgs banget untuk cek antrean	aplikasi jkn ini bagus banget untuk cek antrean

#### 6. Menyimpan Dataset

Tahap terakhir adalah menyimpan hasil dari proses *text preprocessing* ke dalam format CSV menggunakan `pandas.to_csv()` agar dapat digunakan pada tahap analisis berikutnya.

### 3.4.2 *Labeling Data*

Labeling data dilakukan setelah preprocessing data dan prosesnya dilakukan secara manual sesuai dengan aspek, sentimen dan kriteria yang telah ditentukan yaitu untuk *usability* (Kemudahan Penggunaan) ditandai dengan sejauh mana user dapat menggunakan aplikasi tanpa perlu belajar, *Reliability* (Keandalan) ditandai dengan sejauh mana system dapat mempertahankan kinerja dalam kondisi tentu dan dalam jangkauan waktu tertentu, dan *efficiency* ditandai dengan sejauh mana sistem memberikan performa yang mencakup perilaku waktu. Proses dilakukan dengan bantuan label studio versi 1.20, berikut merupakan tahapan proses dari penggunaan label studio.

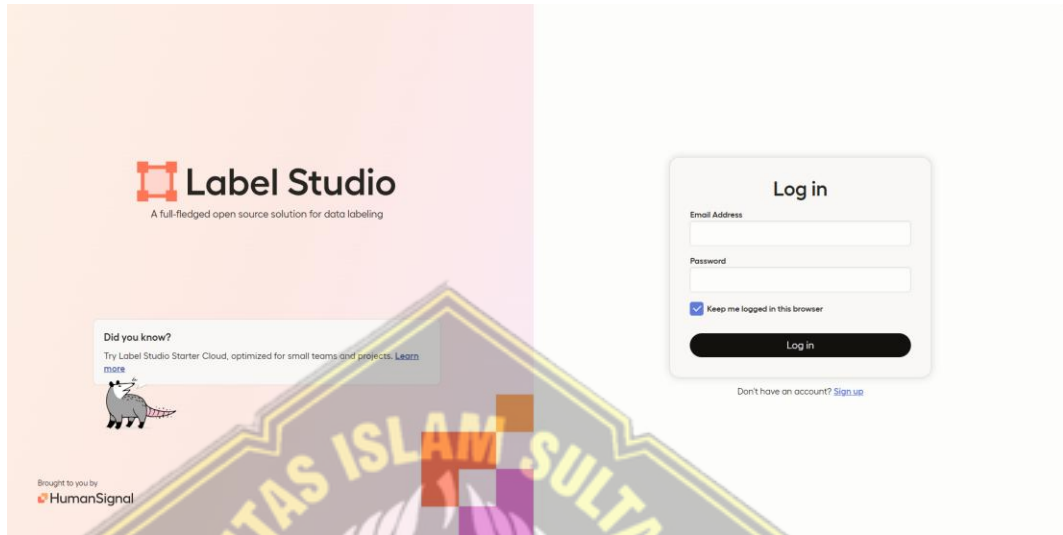
#### 1. *Install* Label Studio

Pada tahap melakukan install label studio dengan perintah `pip install label-studio`”.

#### 2. *Start* Label Studio

Lalu jalankan label studi yang telah *diinstall* dengan perintah “`label-studio start`”

### 3. *Login* label Studio



Gambar 3. 4 Login label studio

Pada gambar 3.4 merupakan tampilan ketika label studio pertama kali diakses, yaitu terdapat form login untuk masuk ke menu utama.

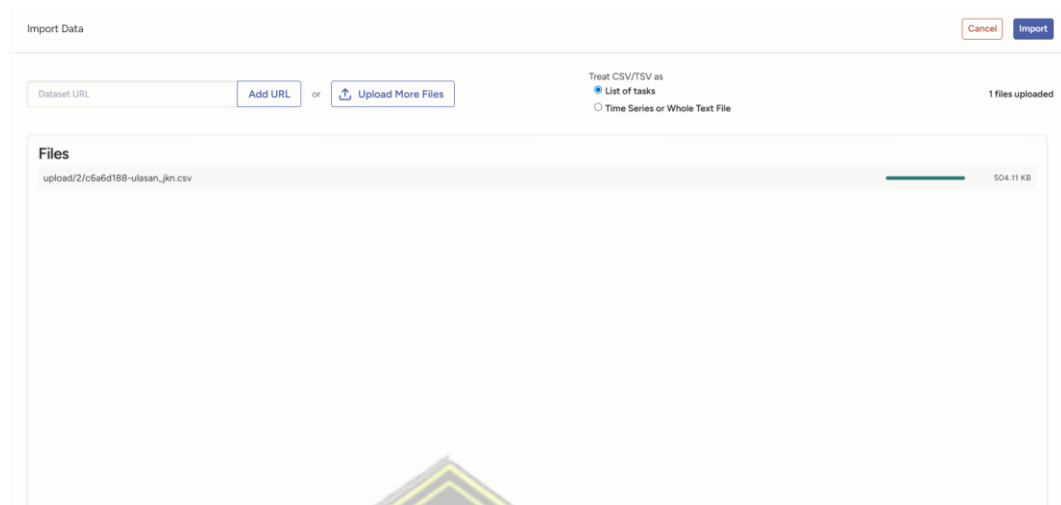
### 4. Membuat *project* label studio



Gambar 3. 5 Membuat *project* label studio

Pada gambar 3.5 merupakan proses membuat *project* pada label studio dengan tujuan untuk melakukan anotasi pada dataset yang diinginkan.

### 5. *Import* dataset



Gambar 3. 6 Import Dataset

Pada gambar 3.6 merupakan tampilan ketika *import* dataset, hal tersebut dilakukan untuk melakukan import pada dataset yang ingin dianotasi.

## 6. Konfigurasi Label Studio

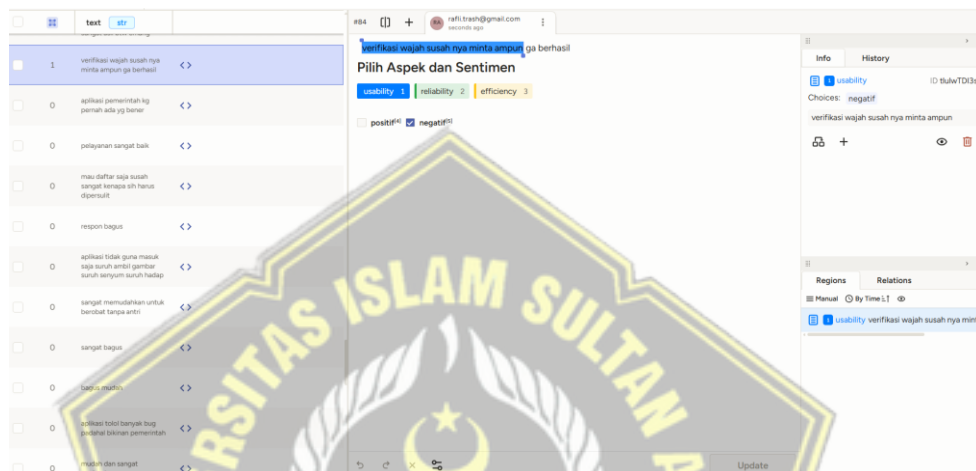


Gambar 3. 7 Konfigurasi Label Studio

Pada gambar 3.7 melakukan konfigurasi label studio, terlihat pada bagian *configure* data merupakan proses menentukan sumber teks yang akan dilabeli. Pada pengaturan ini, digunakan opsi "*Use text from*" dengan nilai "\$text" yang berarti sistem akan mengambil data teks dari *field* yang telah ditentukan sebelumnya dalam dataset. Pada *add label names* berfungsi untuk menentukan aspek-aspek apa saja yang akan digunakan dalam pelabelan. Dalam proyek ini terdapat tiga label aspek yang telah ditambahkan yaitu *usability* dengan

penanda warna biru, *reliability* dengan penanda warna hijau, dan *efficiency* dengan penanda warna kuning. Bagian *add choices* digunakan untuk menentukan pilihan sentimen yang akan diberikan pada setiap aspek yang telah ditandai. Dalam konfigurasi ini terdapat dua choices yaitu "positif" dan "negatif" yang merepresentasikan polaritas sentimen dari teks yang dianalisis.

## 7. Pelabelan Teks



Gambar 3. 8 Pelabelan teks

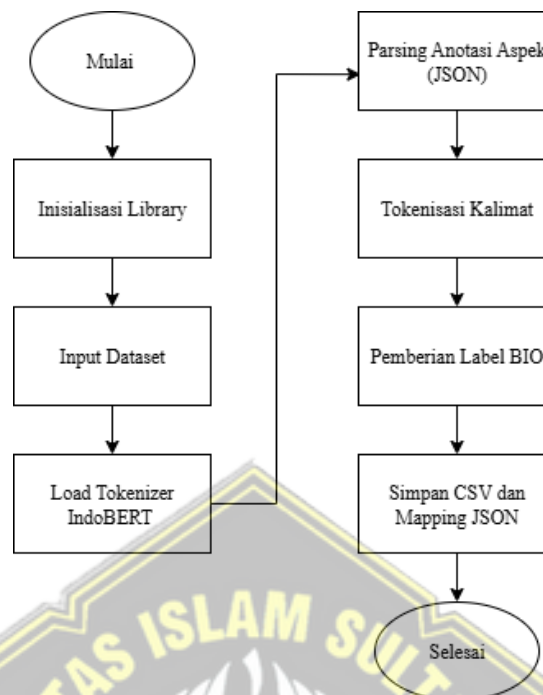
Pada gambar 3.4 merupakan proses pelabelan teks, yang diawali dengan melakukan *highlight* atau menandai bagian teks yang mengandung opini tentang aspek tertentu pada panel tengah. Setelah teks ditandai, pengguna memilih label aspek yang sesuai dengan kriteria dengan cara mengklik salah satu dari pilihan yang tersedia yaitu *usability*, *reliability*, atau *efficiency*. Selanjutnya menentukan sentimen dengan menambahkan tag positif atau negatif sesuai dengan makna dan konteks teks tersebut. Hasil pelabelan kemudian disimpan dengan menekan tombol "*Update*" yang tersedia di bagian bawah *interface*. Proses sistematis ini dilakukan untuk membuat dataset berlabel yang nantinya akan digunakan untuk melatih model analisis sentimen berbasis aspek secara otomatis. Berikut merupakan hasil setelah dataset dilakukan anotasi.

Tabel 3. 5 Labeling data

<i>text</i>	<i>decisions</i>
mau verifikasi wajah pun tidak bisa	{"verifikasi wajah pun tidak bisa": ["reliability", "negatif"]}
untuk memverifikasi wajah sangat lah susah selalu gagal	{"memverifikasi wajah sangat lah susah": ["usability", "positif"], "selalu gagal": ["reliability", "negatif"]}
aplikasi nya sangat membantu	{"aplikasi nya sangat membantu": ["usability", "positif"]}
mau daftar saja susah nya minta ampun aplikasi bukan mempermudah tapi mempersulit	{"mau daftar saja susah nya minta ampun aplikasi bukan mempermudah tapi mempersulit": ["usability", "negatif"]}

### 3.5 *Token classification*

Pada tahap *Token Classification* dalam ABSA, setiap kalimat terlebih dahulu ditokenisasi menggunakan *tokenizer* dari model IndoBERT. Tokenisasi ini memecah kalimat menjadi unit-unit token yang dapat diproses model. Selanjutnya, setiap token hasil tokenisasi disejajarkan dengan label yang telah ditentukan berdasarkan skema anotasi aspek. Dalam sebuah kata atau frasa yang merepresentasikan aspek tertentu diberi label sesuai skema BIO (*Begin, Inside, Outside*), sedangkan token lain yang tidak terkait aspek diberi label *O*. Hasil akhir dari tahap ini adalah dataset siap latih yang terdiri atas pasangan *token* dan *label aspek*, yang selanjutnya digunakan oleh model untuk mempelajari pola kemunculan aspek dalam teks sehingga model mampu melakukan identifikasi aspek secara otomatis dari ulasan, sebagai langkah awal dalam analisis sentimen berbasis aspek. Berikut alur dari *token classification*.



Gambar 3. 9 Alur *token classification*

Pada gambar 3.5 merupakan proses alur dari *token classification* berikut penjelasan dari setiap alur tersebut:

1. Inisialisasi *library*

Pada tahapan inisialisasi *library* dilakukan pemanggilan pustaka Python yang diperlukan. Beberapa *library* yang digunakan yaitu:

a. Pandas

Digunakan untuk membaca dan mengelola dataset dalam format tabel dalam hal ini untuk membaca file CSV.

b. JSON

Digunakan untuk membaca dan memproses data anotasi aspek dalam format JSON.

c. *Transformers*

*Library Transformers* dari *Hugging Face*, digunakan untuk memuat tokenizer IndoBERT agar teks dapat dipecah menjadi token sesuai model yang digunakan.

d. *tqdm*

Digunakan untuk menampilkan *progress bar* selama proses berjalan agar memudahkan pemantauan.

e. Numpy

Digunakan untuk mendukung pengolahan data numerik selama proses pelabelan.

2. Input dataset

Dataset yang digunakan berasal dari hasil anotasi ulasan pengguna tiga aplikasi layanan publik yaitu Mobile JKN, Signal, dan M-Paspor. Setiap file berformat CSV berisi teks ulasan dan hasil anotasi aspek-sentimen yang nantinya akan dipetakan ke setiap token. Tahap ini memastikan data masukan telah terbaca dengan benar sebelum dilakukan pemrosesan lebih lanjut.

3. Load tokenizer IndoBERT

Pada tahap ini dilakukan pemanggilan tokenizer IndoBERT melalui pustaka *transformers*. Tokenizer ini berfungsi untuk memecah setiap kalimat menjadi unit kata (token) yang sesuai dengan format masukan model BERT. Dengan tokenisasi yang konsisten, setiap kata dalam ulasan dapat diproses dengan akurat oleh model bahasa.

4. Parsing anotasi aspek (JSON)

Tahapan ini bertugas membaca file anotasi aspek dalam format JSON yang berisi informasi pasangan kata atau frasa dengan kategori aspek dan sentimennya. Contohnya frasa "*tampilan aplikasinya mudah digunakan*" dapat dianotasi sebagai aspek *Usability* dengan sentimen positif. Proses parsing ini memastikan setiap anotasi dapat dikenali dan digunakan pada tahap pelabelan token.

5. Tokenisasi kalimat

Setiap kalimat pada dataset diubah menjadi kumpulan token menggunakan tokenizer IndoBERT. Tokenisasi ini membantu sistem memetakan setiap bagian teks secara lebih terstruktur sehingga lebih mudah diberikan label aspek yang sesuai

Tabel 3. 6 Tokenisasi teks

<i>Before</i>	<i>After</i>
tidak berguna verifikasi wajah dari pagi malam pagi ulang tetap gagal terus	['tidak', 'berguna', 'verifikasi', 'wajah', 'dari', 'pagi', 'malam', 'pagi', 'ulang', 'tetap', 'gagal', 'terus']

## 6. Pemberian label BIO

Setelah kalimat berhasil ditokenisasi, setiap token diberi label menggunakan skema BIO (*Begin-Inside-Outside*). Label ini menandai token yang termasuk ke dalam bagian aspek tertentu, seperti:

- a. B untuk awal frasa positif atau negatif,
- b. I untuk token berikutnya dalam frasa yang sama,
- c. O untuk token yang tidak berkaitan dengan aspek apa pun.

Skema ini membantu model memahami batasan antar aspek dalam kalimat secara lebih jelas.

Tabel 3. 7 label BIO

<i>Before</i>	<i>After</i>
['tidak', 'berguna', 'verifikasi', 'wajah', 'dari', 'pagi', 'malam', 'pagi', 'ulang', 'tetap', 'gagal', 'terus']	['O', 'O', 'B-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif', 'I-efficiency-negatif']

## 7. Simpan CSV dan Mapping JSON

Pada tahapan terakhir Hasil dari proses pelabelan disimpan dalam dua format:

- a. File CSV yang berisi teks, token, dan label BIO yang sudah siap digunakan pada tahap pelatihan model

- b. File JSON, yang berisi peta hubungan antara label dan ID (label2id serta id2label), digunakan untuk menjaga konsistensi label saat pelatihan dan evaluasi.

Dengan alur ini, dataset mentah yang berisi ulasan dan anotasi aspek berhasil diproses menjadi format terstruktur yang siap digunakan dalam pembangunan sistem *Aspect-Based Sentiment Analysis* (ABSA) berbasis IndoBERT.

### 3.6 *Fine-tuning* IndoBERT

Model IndoBERT melalui proses *fine-tuning* dengan proses pelatihan dilakukan dengan *hyperparameter* Epoch 5. Jumlah *epoch* ini dipilih berdasarkan observasi pada data validasi, di mana performa model cenderung mencapai titik optimal dan mulai stagnan setelah *epoch* ke-5. *Learning Rate*: 3e-5 nilai ini merupakan laju belajar yang umum direkomendasikan untuk proses *fine-tuning* model berbasis BERT dan terbukti memberikan hasil yang stabil. Batch Size 8 ukuran *batch* ini dipilih sebagai kompromi antara kecepatan pelatihan dan keterbatasan memori komputasi (GPU VRAM). Pelatihan diimplementasikan menggunakan *library HuggingFace Transformers*. berikut merupakan tabel *hyperparameter* yang digunakan.

Tabel 3. 8 Parameter *fine-tuning*

<i>Hyperparameter</i>	Nilai	<i>description</i>
<i>Max Sequence Length</i>	256	Panjang maksimum token dalam satu input teks. Jika teks melebihi batas ini, sisanya akan dipotong (truncated).
<i>Batch Size</i>	8	Jumlah sampel yang diproses sekaligus dalam satu iterasi sebelum memperbarui bobot model.
<i>Epoch</i>	5	Jumlah kali seluruh dataset dilalui oleh model selama proses pelatihan. Semakin

<i>Hyperparameter</i>	<i>Nilai</i>	<i>description</i>
		banyak epoch, semakin lama training dan potensi overfitting meningkat.
<i>Learning Rate</i>	3e-5 (0.00003)	Kecepatan model dalam memperbarui bobot selama training. Nilai kecil membuat pembelajaran lebih stabil tetapi lebih lambat.
<i>Weight Decay</i>	0.01	Nilai regularisasi yang membantu mencegah <i>overfitting</i> dengan mengecilkan bobot besar secara bertahap selama training.

### 3.7 Evaluasi Model

Evaluasi akhir model dilakukan pada data uji yang belum pernah dilihat oleh model sebelumnya. Performa model diukur menggunakan metrik standar untuk tugas *sequence labeling* dari pustaka sequeval, yang meliputi *accuracy precision*, *recall*, *f1-score* perhitungan metrik dilakukan secara mikro (*micro average*) dan makro (*macro average*) untuk mendapatkan gambaran performa model secara keseluruhan dan per kategori label.

### 3.8 Analisis Hasil

Pada analisis hasil merupakan langkah untuk mengidentifikasi dan membahas temuan yang diperoleh dari proses evaluasi model. Analisis ini bertujuan untuk menjawab rumusan masalah yang telah ditetapkan. Rencana analisis hasil akan difokuskan pada analisis performa model dengan menganalisis hasil metrik evaluasi *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari ketiga model yang telah dilatih pada dataset yang bersumber dari ulasan Mobile JKN, M-Paspor, dan SIGNAL di *google play store*. Analisis ini juga mencakup perbandingan performa antar model untuk

mengidentifikasi model mana yang paling efektif dan faktor-faktor yang memengaruhinya seperti kualitas dan kuantitas data. Selain itu akan dilakukan pula analisis aspek dominan untuk mengidentifikasi dan menganalisis aspek *Usability*, *Reliability*, *Efficiency* yang paling dominan menerima sentimen positif dan negatif pada setiap aplikasi. Analisis aspek ini dilakukan berdasarkan hasil prediksi model pada data uji untuk mendapatkan wawasan kualitatif mengenai keluhan utama dan keunggulan utama dari setiap aplikasi.

### 3.9 Implementasi Sistem

#### 3.9.1 Perancangan *interface*

Pada tahap perancangan *interface*, dilakukan proses desain dan pembuatan rancangan visual yang akan menjadi jembatan interaksi antara pengguna (*user*) dengan sistem.



Gambar 3. 10 Halaman utama

Gambar 3.3 merupakan tampilan *interface* ketika pertama kali diakses. Pada halaman ini, pengguna dapat melihat dua komponen input utama yang dirancang untuk memudahkan interaksi dengan sistem. Komponen pertama adalah sebuah

dropdown menu yang memungkinkan pengguna untuk memilih model klasifikasi yang diinginkan dari beberapa opsi model yang telah tersedia. Komponen kedua adalah sebuah text area input yang didesain khusus untuk memungkinkan pengguna memasukkan teks ulasan yang akan dianalisis oleh sistem.

### 3.9.2 Tahapan Kerja Sistem

Pembuatan sistem dilakukan dengan meimplementasi dari sistem *Aspect-Based Sentiment Analysis* (ABSA) yang sebelumnya telah dirancang. Sistem ini dibangun menggunakan *framework Streamlit* dan dihubungkan melalui file utama `app.py`, yang di dalamnya telah dimuat model IndoBERT-ABSA yang telah dilatih. Tahapan kerja sistem secara umum dapat dilihat pada Gambar 3.11.



Gambar 3. 11 Tahapan kerja pembuatan sistem

Berikut ini penjelasan dari masing-masing tahapan berdasarkan alur sistem yang telah diterapkan pada antarmuka *Streamlit*:

#### 1. Mulai

Sistem akan aktif dan siap digunakan ketika pengguna membuka aplikasi melalui antarmuka *Streamlit*. Proses ini merupakan awal dari interaksi antara pengguna dengan sistem analisis sentimen berbasis aspek.

#### 2. Input Data

Pada tahap input data, *user* memasukkan teks ulasan yang ingin dianalisis. Setelah data dimasukkan, sistem akan secara otomatis menjalankan proses pra-pemrosesan, meliputi ,pembersihan teks dari karakter tidak relevan, tokenisasi untuk memecah kalimat menjadi bagian-bagian kecil, normalisasi untuk menyeragamkan format teks.

#### 3. Prediksi dan Hasil Analisis

Setelah data diproses, sistem akan melakukan prediksi aspek dan sentimen. Hasilnya ditampilkan.

#### 4. Selesai

Setelah proses selesai, pengguna dapat kembali ke tahap input untuk menganalisis data lain jika diperlukan.



## BAB IV

### HASIL DAN ANALISIS PENELITIAN

#### 4.1 Persiapan Data

Data dalam penelitian ini dikumpulkan menggunakan teknik *web scraping* dari tiga aplikasi layanan publik yang tersedia di Google Play Store, yaitu Mobile JKN, M-Paspor, dan SIGNAL. Masing-masing aplikasi diambil sebanyak 1.700 data ulasan pengguna sehingga total data yang terkumpul berjumlah 5.100 ulasan sebelum dilakukan *preprocessing*. Proses pengambilan data dilakukan dengan bantuan pustaka *google-play-scraper* menggunakan bahasa pemrograman Python. Setiap aplikasi diidentifikasi menggunakan application ID yang menjadi parameter utama dalam proses *scraping*, misalnya *app.bpjs.mobile* untuk aplikasi Mobile JKN. Ulasan yang diambil difokuskan pada pengguna berbahasa Indonesia dengan lokasi di wilayah Indonesia. Untuk menghindari adanya pembatasan akses dari pihak Google, pengambilan data dilakukan secara bertahap (*batching*) dengan memberikan jeda waktu antar permintaan data. Seluruh hasil *scraping* kemudian disimpan dalam format CSV agar lebih mudah diproses pada tahap selanjutnya.

Hasil *scraping* menghasilkan beberapa kolom yang membentuk struktur awal dataset, yaitu No, reviewId, userName, userImage, content, dan score. Kolom No berfungsi sebagai nomor urut data untuk mempermudah identifikasi, sedangkan reviewId merupakan kode unik yang diberikan oleh Google Play pada setiap ulasan. Kolom userName menyimpan nama pengguna yang menuliskan ulasan dan userImage berupa tautan menuju foto profil pengguna di Google Play. Kolom yang paling penting adalah content karena berisi teks ulasan yang akan dianalisis menggunakan metode *Aspect-Based Sentiment Analysis (ABSA)*, sementara kolom score berisi nilai rating dalam skala 1 hingga 5 yang diberikan oleh pengguna terhadap aplikasi. Rating ini dapat digunakan sebagai informasi tambahan untuk memvalidasi kecenderungan sentimen positif atau negatif yang muncul dalam teks ulasan. Dengan struktur data tersebut, penelitian dapat memfokuskan analisis pada konten ulasan pengguna sebagai bahan utama untuk mendeteksi aspek serta

sentimen, sementara atribut lain seperti reviewId, userName, dan userImage hanya bersifat sebagai metadata pendukung.

Tabel 4. 1 Hasil *Web Scraping*

No	reviewId	userName	userImage	content	score
1.	3386d9a6-772f-421e-a5d1-4207d8379cf7	Joni Maulana	https://play-lh.googleusercontent.com/a/ACg8ocJ9pIk9ygyHF3_WqmXgh-f72tkIrc-pG_xudyMNfGo7w5gw=mo	Ada nya apk ini bukan diper mudah,malah dipersulit.	1
2.	be410b36-afa1-499e-b266-dbef5f829a8c	Andi Susanto	https://play-lh.googleusercontent.com/a/ACg8ocKMneAmnZ14yMBoItDX0Hmefmox1YK6j57ZbekFJGA0e6ZUp28=mo	Tingkatkan kecepatan kode Verifikasi no Hp n email , bisa ubah faskes email dan no HP diaplikasi, tanpa ke kantor BPJS. Terima kasih	5
...				....	
1500.	a6a4ab8a-0b99-4381-a207-	choirun najwa	https://play-lh.googleusercontent.com/	aplikasi sering ga sinkron sama	3

No	reviewId	userName	userImage	content	score
	5c74f35a3b84		a-/ALV-UjVu16xmNZoH1wNHFnJjiBIajF-1kFKcbCKH_wZIIIn4QZA hVtKNKcg	agak lama, padahal sering di update trs di refresh juga	

## 4.2 Pra-Pemrosesan Data

### 4.2.1 Text Preprocessing

Untuk menyiapkan data teks ulasan sebelum pelabelan dan pelatihan model, dilakukan proses pra-pemrosesan data. Proses ini melakukan beberapa tahapan sebagai berikut:

#### 1. Case Folding

Tabel 4. 2 Case Folding

Aplikasi	Sebelum Case Folding	Setelah Case Folding
Mobile JKN	Mau cek status kepesertaan jkn ribet banget gagal trs ...coba hubungi pandawa gk bisa membantu tmbh muter"tolong dong dipermudah jmo mobile bagus gk ada masah ribet jkn mobile 😊😊😊😊😊😊	mau cek status kepesertaan jkn ribet banget gagal trs ...coba hubungi pandawa gk bisa membantu tmbh muter"tolong dong dipermudah jmo mobile bagus gk ada masah ribet jkn mobile 😊😊😊😊😊😊
SIGNAL	Bayar pajk online ribett, sering err0r trs aplikasiny force close mluuuu	bayar pajk online ribett, sering err0r trs aplikasiny force close mluuuu

Aplikasi	Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
	😬😬 susah bgt masukin plat nomerr padahal udh bnerrr, bikin emosi	😬😬 susah bgt masukin plat nomerr padahal udh bnerrr, bikin emosi
M-Paspor	Proses daftar paspor online ribettt server down mlu... 😬😬 antriannya ga jelas trus harus ulang2 upload dokumen, pusing banget sumpahhh	proses daftar paspor online ribettt server down mlu... 😬😬 antriannya ga jelas trus harus ulang2 upload dokumen, pusing banget sumpahhh

Pada tabel 4.2 dilakukan proses case folding, yaitu mengubah seluruh teks menjadi huruf kecil. Langkah ini bertujuan untuk menjaga konsistensi data sehingga kata dengan bentuk penulisan berbeda, misalnya “BPJS”, “bpjs”, atau “Bpjs” tetap dianggap sebagai satu kata yang sama. Dengan demikian, variasi penulisan huruf besar dan kecil tidak memengaruhi proses tokenisasi maupun hasil analisis, sehingga model dapat bekerja lebih optimal.

## 2. *Cleaning Data*

Tabel 4. 3 Data Sebelum dan Sesudah *Cleaning* pada Ulasan Aplikasi

Aplikasi	Sebelum <i>Cleaning Data</i>	Setelah <i>Cleaning Data</i>
Mobile JKN	mau cek status kepesertaan jkn ribet banget gagal trs ...coba hubungi pandawa gk bisa membatu tmbh muter"tolong dong dipermudah jmo mobile bagus gk ada masah ribet jkn mobile 😬😬😬😬😬😬	mau cek status kepesertaan jkn ribet banget gagal trs coba hubungi pandawa gk bisa membatu tmbh muter tolong dong dipermudah jmo mobile bagus gk ada masah ribet jkn mobile

Aplikasi	Sebelum <i>Cleaning Data</i>	Setelah <i>Cleaning Data</i>
SIGNAL	bayar pajk online ribett, sering err0r trs aplikasiny force close mluuuu 🙄🙄 susah bgt masukin plat nomerr padahal udh bnerrr, bikin emosi	bayar pajk online ribett sering err0r trs aplikasiny force close mluuuu susah bgt masukin plat nomerr padahal udh bnerrr bikin emosi
M-Paspor	proses daftar paspor online ribettt server down mlu... 🙄🙄 antriannya ga jelas trus harus ulang2 upload dokumen, pusing banget sumpahhh	proses daftar paspor online ribettt server down mlu antriannya ga jelas trus harus ulang2 upload dokumen pusing banget sumpahhh

Pembersihan data pada Tabel 4.3 dilakukan untuk mengurangi *noise* dan memastikan teks ulasan siap diproses oleh model. Tahapan ini mencakup penghapusan karakter non-ASCII seperti emotikon, simbol khusus, dan huruf dari alfabet non-Latin agar data hanya berisi karakter standar. Selain itu, tautan atau URL yang tidak relevan terhadap analisis juga dihapus karena hanya menambah gangguan pada teks. Tanda baca yang berlebihan seperti penggunaan berulang tanda seru atau tanda tanya turut dihilangkan untuk mencegah bias pada model. Angka yang tidak memiliki makna kontekstual juga dibuang, kecuali jika memiliki peran penting dalam teks, misalnya nomor antrean. Selanjutnya, normalisasi spasi dilakukan dengan menghapus spasi ganda maupun spasi di awal dan akhir kalimat sehingga teks lebih rapi dan konsisten.

### 3. Normalization

Tabel 4. 4 Normalization

<b>Aplikasi</b>	<b>Sebelum Normalization</b>	<b>Setelah Normalization</b>
Mobile JKN	mau cek status kepesertaan jkn ribet sekali sering gagal trs coba hubungi pandawa ga bisa membantu malah tmbh muter tolong dipermudah jmo mobile bagus gak ada masalah gak ada masalah ribet jkn mobil	mau cek status kepesertaan jkn ribet banget gagal terus coba hubungi pandawa tidak bisa membantu malah tambah muter tolong dipermudah jkn mobile bagus tidak ada masalah ribet jkn mobile
SIGNAL	bayar pajak online ribet sering error terus aplikasinya force close melulu susah banget masukin plat nomor padahal sudah benar bikin emosi	bayar pajak online ribet sering error terus aplikasinya force close melulu susah banget memasukkan plat nomor padahal sudah benar bikin emosi
M-Paspor	proses daftar paspor online ribet server down melulu antriannya tidak jelas harus ulang ulang upload dokumen pusing banget.	proses daftar paspor online ribet server down melulu antriannya tidak jelas terus harus ulang upload dokumen pusing banget sumpah

Pada Tabel 4.4 dilakukan proses *normalization*, yaitu mengubah kata tidak baku, singkatan, atau penulisan tidak standar menjadi bentuk kata yang sesuai dengan kaidah bahasa Indonesia. Tahapan ini penting untuk menjaga konsistensi teks sehingga kata yang berbeda penulisannya tetapi memiliki makna sama dapat diperlakukan sebagai satu representasi. Contohnya, kata

“trs” diubah menjadi “terus”, “ga” menjadi “tidak”, dan “tmbh” menjadi “tambah”. Dengan adanya normalisasi, kualitas data menjadi lebih baik dan model dapat memahami teks secara lebih akurat tanpa terpengaruh variasi penulisan.

#### 4.2.2 Labelling data

Proses pelabelan data dilakukan secara manual menggunakan Label Studio, sebuah alat anotasi *open source*. Dari proses ini didapatkan total 4.500 data ulasan yang berasal dari tiga aplikasi Mobile JKN (1500 data), SIGNAL (1500 data), dan M-Paspor (1500 data). Setiap ulasan tersebut ditinjau dan diberi label berdasarkan tiga aspek yaitu *usability*, *reliability*, dan *efficiency*, beserta sentimennya positif atau negatif. Sebanyak 600 data tidak disertakan karena tidak mengandung kata kunci atau pembahasan yang masuk ke dalam salah satu dari tiga aspek yang diteliti. Selama proses pelabelan, ditemukan bahwa sebagian besar ulasan hanya memuat satu aspek (*single-aspect*), namun terdapat proporsi yang cukup signifikan M-Paspor, SIGNAL, dan Mobile JKN yang mengandung lebih dari satu aspek (*multi-aspect*). Tujuan dari pelabelan ini adalah untuk mengidentifikasi polaritas sentimen pada setiap aspek sehingga hasilnya dapat digunakan sebagai data latih untuk pemodelan analisis sentimen berbasis aspek. Pada table 4.6 merupakan contoh hasil dari labeling data.

Tabel 4. 5 Tabel Hasil Anotasi Data

No.	Aplikasi	Ulasan	Notasi
1.	Mobile JKN	aplikasinya kurang responsif kendala di scan wajah verifikasi sms yg super lama ganti nomor tidak pernah ada verifikasi sms ayo lah orang it nya di buat aplikasi agar lebih lancar	{"aplikasinya kurang responsif": ["reliability", "negatif"], "kendala di scan wajah verifikasi sms yg super lama": ["efficiency", "negatif"], "ganti nomor tidak pernah ada verifikasi

No.	Aplikasi	Ulasan	Notasi
			sms": ["reliability", "negatif"]}
2.	SIGNAL	aplikasinya gampang dipakai dan hemat waktu tapi sering banget tertutup sendiri	{"aplikasinya gampang dipakai": ["usability", "positif"], "hemat waktu": ["efficiency", "positif"], "sering banget tertutup sendiri ": ["reliability", "negatif"]}"
3.	M-Paspor	aplikasi nya tidak banget ini susah banget login dan isi data sering keluar sendiri	{"aplikasi nya tidak banget ini susah banget login": ["usability", "negatif"], "isi data sering keluar sendiri": ["reliability", "negatif"]}"

Pada tabel 4.5 menunjukkan hasil proses anotasi ulasan pengguna aplikasi dengan menggunakan pendekatan *Aspect-Based Sentiment Analysis (ABSA)*. Setiap ulasan dipetakan ke dalam pasangan aspek dan sentimen dalam format *key-value* JSON, di mana *key* merupakan potongan kalimat yang merepresentasikan opini pengguna, sedangkan *value* berupa pasangan label yang terdiri atas kategori aspek (*usability*, *reliability*, *efficiency*) serta sentimen (positif atau negatif).

Sebagai contoh, pada aplikasi Mobile JKN, opini "aplikasinya kurang responsif" dikategorikan ke aspek *reliability* dengan sentimen negatif, sementara "kendala di scan wajah verifikasi sms yg super lama" masuk ke aspek *efficiency* dengan sentimen negatif. Hal serupa juga terlihat pada aplikasi SIGNAL, di mana ulasan positif seperti "aplikasinya gampang dipakai" dan "hemat waktu" dipetakan

ke aspek *usability* dan *efficiency* dengan sentimen positif, sedangkan keluhan "sering banget tertutup sendiri" masuk ke aspek *reliability* dengan sentimen negatif. Adapun pada aplikasi M-Paspor, ulasan negatif seperti "susah banget login" dan "isi data sering keluar sendiri" juga diberi label sesuai aspek yang relevan. Setelah proses labeling dilakukan, tahap berikutnya adalah menghitung distribusi ulasan berdasarkan aspek dan sentimen pada masing-masing aplikasi:

Tabel 4. 6 Distribusi aspek dan sentimen Mobile JKN

Mobile JKN			
Aspek	Negatif	Positif	Total per Aspek
<i>Usability</i>	341	511	852
<i>Reliability</i>	458	158	616
<i>Efficiency</i>	184	145	329

Tabel 4. 7 Distribusi aspek dan sentimen M-Paspor

M-Paspor			
Aspek	Negatif	Positif	Total per Aspek
<i>Usability</i>	407	247	654
<i>Reliability</i>	833	212	1045
<i>Efficiency</i>	171	208	379

Tabel 4. 8 Distribusi aspek dan sentimen SIGNAL

SIGNAL			
Aspek	Negatif	Positif	Total per Aspek
<i>Usability</i>	170	791	961
<i>Reliability</i>	294	225	519
<i>Efficiency</i>	127	410	537

Hasil perhitungan pada Tabel 4.7 menunjukkan bahwa pada aplikasi Mobile JKN, aspek *usability* mendominasi dengan total 852 ulasan, terdiri dari 341 ulasan negatif dan 511 ulasan positif. Aspek *reliability* menempati posisi kedua dengan 616 ulasan (458 negatif dan 158 positif), sementara *efficiency* memiliki jumlah terendah yaitu 329 ulasan (184 negatif dan 145 positif).

Pada Tabel 4.8, distribusi ulasan aplikasi M-Paspor memperlihatkan bahwa aspek *reliability* memiliki jumlah ulasan tertinggi, yakni 1.045 ulasan (833 negatif

dan 212 positif). Aspek *usability* menyusul dengan 654 ulasan (407 negatif dan 247 positif), sedangkan *efficiency* memiliki 379 ulasan (171 negatif dan 208 positif).

Selanjutnya, pada Tabel 4.9 terlihat bahwa aplikasi SIGNAL cenderung didominasi oleh aspek *usability* dengan total 961 ulasan, mayoritas bersentimen positif (791 positif, 170 negatif). Aspek *efficiency* menempati urutan kedua dengan 537 ulasan (410 positif dan 127 negatif), sementara *reliability* memiliki 519 ulasan (294 negatif dan 225 positif).

Hasil distribusi pada ketiga aplikasi menunjukkan bahwa aspek *usability* dan *reliability* lebih sering muncul dalam ulasan dibandingkan *efficiency*. Selain itu, kecenderungan sentimen juga berbeda pada tiap aplikasi: Mobile JKN dan M-Paspor lebih banyak menerima ulasan negatif pada aspek *reliability*, sedangkan SIGNAL mendapatkan dominasi ulasan positif terutama pada aspek *usability* dan *efficiency*.

#### 4.2.3 *Balancing*

Dataset awal mengalami ketidakseimbangan jumlah ulasan antar kelas aspek-sentimen, yang berpotensi menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas. Kompleksitas masalah ini meningkat karena struktur data bersifat multi-aspek, di mana satu ulasan dapat memiliki lebih dari satu label aspek-sentimen. Untuk mengatasinya, diterapkan strategi penyeimbangan data menggunakan metode *oversampling* yang dirancang untuk mempertahankan struktur multi-label tersebut.

Proses *balancing* ini dilakukan secara spesifik per aplikasi dan menargetkan level kombinasi aspek-sentimen, seperti *usability positif* atau *reliability negatif*, bukan hanya aspek atau sentimen secara terpisah. Dengan melakukan skema *oversampling* ulasan dari kelas-kelas minoritas hingga jumlahnya setara dengan kelas mayoritas pada setiap ulasan. Dalam proses ini ulasan multi-aspek tetap dipertahankan secara utuh, sehingga satu ulasan dapat diduplikasi jika mengandung label minoritas. Pendekatan ini memastikan bahwa model tidak hanya belajar dari data yang lebih seimbang, tetapi juga dari variasi konteks yang lebih kaya yang terdapat dalam ulasan multi-aspek, sehingga dapat mengurangi bias dan

meningkatkan kemampuan generalisasi model. Distribusi data setelah balancing dapat dilihat pada table 4.11

Tabel 4. 9 Jumlah data Mobile JKN sebelum dan setelah *balancing*

<b>Mobile JKN</b>		
<b>Label</b>	<b>Sebelum <i>Balancing</i></b>	<b>Setelah <i>Balancing</i></b>
<i>Usability</i> - positif	511	879
<i>Reliability</i> - positif	158	827
<i>Efficiency</i> - positif	145	870
<i>Usability</i> - negatif	341	831
<i>Reliability</i> - negatif	458	907
<i>Efficiency</i> - negatif	184	645

Tabel 4. 10 Jumlah data M-Paspor sebelum dan setelah *balancing*

<b>M-Paspor</b>		
<b>Label</b>	<b>Sebelum <i>Balancing</i></b>	<b>Setelah <i>Balancing</i></b>
<i>Usability</i> - positif	247	1713
<i>Reliability</i> - positif	212	1602
<i>Efficiency</i> - positif	208	1607
<i>Usability</i> - negatif	407	1373
<i>Reliability</i> - negatif	833	1908
<i>Efficiency</i> - negatif	171	1095

Tabel 4. 11 Jumlah data SIGNAL sebelum dan setelah *balancing*

<b>SIGNAL</b>		
<b>Label</b>	<b>Sebelum <i>Balancing</i></b>	<b>Setelah <i>Balancing</i></b>
<i>Usability</i> - positif	791	1786
<i>Reliability</i> - positif	225	1260
<i>Efficiency</i> - positif	410	1398
<i>Usability</i> - negatif	170	1153
<i>Reliability</i> - negatif	294	1145
<i>Efficiency</i> - negatif	127	931

Hasil proses *balancing* menunjukkan perubahan distribusi data yang signifikan seperti yang terlihat pada Tabel 4.10, 4.11, dan 4.12. Pada dataset Mobile JKN, kelas yang awalnya sangat minoritas seperti *Efficiency*-positif (145 sampel) dan *Efficiency*-negatif (184 sampel) mengalami peningkatan hingga mendekati jumlah kelas mayoritas. Pola yang sama terlihat pada dataset M-Paspor dimana kelas *Usability*-positif meningkat dari 247 menjadi 1713 sampel, dan pada dataset SIGNAL dimana *Reliability*-positif bertambah dari 225 menjadi 1260 sampel.

Peningkatan jumlah sampel ini tidak hanya menyeimbangkan distribusi kelas, tetapi juga memperkaya variasi linguistik dan kontekstual yang dapat dipelajari oleh model. Dengan jumlah sampel yang lebih representatif untuk setiap kelas, model diharapkan dapat mempelajari pola-pola spesifik yang membedakan setiap kombinasi aspek-sentimen dengan lebih baik, sehingga meningkatkan performa klasifikasi secara keseluruhan. Selain itu, preservasi struktur multi-aspek selama proses *oversampling* memastikan bahwa hubungan antar label dalam satu ulasan tetap terjaga, yang sangat penting untuk mempertahankan konteks natural dari ulasan pengguna. Distribusi data setelah *balancing* menunjukkan bahwa setiap kelas kini memiliki representasi yang cukup untuk meminimalisir bias klasifikasi, sementara tetap mempertahankan karakteristik alami data ulasan yang multi-label.

### 4.3 Token Classification

Pada tahap *token classification* mempersiapkan data dalam format yang sesuai untuk pelatihan model klasifikasi token berbasis aspek. Tahap ini bertujuan untuk mengonversi data ke format token. tokenisasi dilakukan menggunakan tokenizer dari model IndoBERT (*indobenchmark/indobert-base-p1*). Setiap teks ulasan dikonversi ke format token sesuai aspek dan polaritas sentimennya. Pada tabel 4.13 merupakan hasil dari *token classification*.

Tabel 4. 12 Contoh *Token Classification* pada Data Ulasan

APK	Ulasan	Notasi	Token	Label
JKN	melihat rincian denda pelayanan jadi sangat cepat	{"sangat cepat": ["efficiency", "positif"],	['melihat', 'rincian', 'denda',	['O', 'O', 'O', 'O', 'O', 'B-efficiency-



APK	Ulasan	Notasi	Token	Label
			'aplikasi', 'nya', 'malah', 'kaya', 'jaman', 'batu']	negatif, 'I- usability- negatif, 'I- usability- negatif, 'I- usability- negatif, 'I- usability- negatif, 'I- usability- negatif, 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']
SIGNA L	mudah cepat prosesnya jadi ga ada alasan untuk telat bayar pajak	{"mudah": ["usability", "positif"], "cepat prosesnya jadi ga ada alasan untuk telat bayar pajak": ["efficiency", "positif"]}	['mudah', 'cepat', 'prosesnya', 'jadi', 'ga', 'ada', 'alasan', 'untuk', 'telat', 'bayar', 'pajak']	['B- usability- positif, 'B- efficiency- positif, 'I- efficiency- positif, 'I- efficiency- positif, 'I- efficiency-

APK	Ulasan	Notasi	Token	Label
				positif, 'I- efficiency- positif, 'I- efficiency- positif, 'I- efficiency- positif, 'I- efficiency- positif']

Pada tabel 4.13 menampilkan hasil *token classification* dari ulasan pengguna pada tiga aplikasi layanan publik digital, yaitu Mobile JKN, M-Paspor, dan SIGNAL, yang digunakan sebagai sampel dalam penelitian ini. Setiap ulasan dianalisis melalui beberapa tahap mulai dari ekstraksi aspek, penentuan sentimen, hingga proses tokenisasi dan pelabelan dengan skema BIO (*Begin, Inside, Outside*). Kolom *APK* menunjukkan aplikasi asal dari ulasan, misalnya JKN, M-Paspor, atau SIGNAL. Kolom *Ulasan* berisi teks asli yang ditulis pengguna, yang kemudian menjadi bahan utama untuk analisis aspek dan sentimen. Pada kolom *Notasi* ditampilkan representasi pasangan aspek dan polaritas sentimen dalam bentuk frasa yang teridentifikasi, misalnya "sangat cepat" yang dikategorikan sebagai *efficiency-positif* atau "datang langsung harus daftar online..." yang termasuk ke dalam *usability-negatif*. Kolom *Token* menyajikan hasil tokenisasi dari teks ulasan, di mana kalimat dipecah menjadi unit kata per kata agar lebih mudah diproses oleh model. Terakhir, kolom *Label* memperlihatkan anotasi BIO yang menghubungkan setiap token dengan aspek serta sentimen yang sesuai, misalnya token "sangat" dan "cepat" diberi label *B-efficiency-positif* dan *I-efficiency-positif*, sementara token lain yang tidak berhubungan dengan aspek ditandai dengan label *O*. Dengan representasi seperti ini, setiap token dapat dipetakan secara jelas ke aspek dan polaritas tertentu sehingga model mampu mempelajari pola keterkaitan kata, konteks kalimat, dan makna yang lebih mendalam. Proses *token classification* ini sangat penting karena menjadi jembatan antara teks ulasan mentah dengan data

terstruktur yang siap digunakan untuk melatih model *Aspect-Based Sentiment Analysis* berbasis IndoBERT.

#### 4.4 *Fine-tuning* IndoBERT

##### 4.4.1 *Splitting* Data

Pembagian data dilakukan dengan rasio 80:20 yang merupakan salah satu standar pembagian umum. Dengan pembagian data untuk training (80%), validasi (10%), dan *testing* (10%). Pembagian ini bertujuan agar model bisa belajar pola dengan cukup data sehingga nanti bisa memberikan gambaran mengenai kemampuan model dalam menggeneralisasi data baru. Pada Tabel 4.13 merupakan rincian pembagian sampel.

Tabel 4. 13 *Splitting* data peraplikasi

<b>Aplikasi</b>	<b>Data Training</b>	<b>Data Testing</b>	<b>Data Validation</b>
<b>Mobile JKN</b>	2452	307	307
<b>M-Paspor</b>	3998	500	500
<b>SIGNAL</b>	3796	475	475

##### 4.4.2 *Paramater*

Untuk mencapai performa model yang optimal, hyperparameter perlu diatur secara cermat. Pengaturan ini disesuaikan dengan mempertimbangkan spesifikasi perangkat keras dan karakteristik data yang digunakan, karena akan memengaruhi efisiensi komputasi serta akurasi model selama proses pelatihan. Detail konfigurasi hyperparameter yang diterapkan dapat dilihat pada Tabel 4.16

Tabel 4. 14 *Hyperparameter*

<b><i>Hyperparameter</i></b>	<b>Nilai</b>
<i>Max Sequence Length</i>	256
<i>Batch Size</i>	8
<i>Epoch</i>	5
<i>Learning Rate</i>	3e-5 (0.00003)
<i>Weight Decay</i>	0.01

Pemilihan *hyperparameter* dalam penelitian ini disesuaikan dengan kebutuhan model dan keterbatasan perangkat. Nilai *max sequence length* diatur ke

256 untuk memastikan model dapat memproses konteks teks yang relevan, sementara *batch size* 8 dipilih untuk mengoptimalkan penggunaan memori pada perangkat keras yang tersedia. *Learning rate* ditetapkan pada  $3e-5$ , sebuah nilai standar yang memungkinkan proses *fine-tuning* berjalan stabil tanpa merusak pengetahuan dari *pre-trained model*. Selain itu, *weight decay* sebesar 0.01 diterapkan sebagai teknik regularisasi untuk mengurangi risiko *overfitting*. Terakhir, model dilatih selama 8 epoch, yang dinilai cukup bagi model untuk mempelajari pola data secara efektif tanpa membuatnya terlalu spesifik pada data latih (*overfitting*).

#### 4.5 Evaluasi Model

Pada evaluasi model menyajikan hasil akhir dari pelatihan model menggunakan indobert pada tiga dataset ulasan aplikasi yang berbeda. Performa model akan ditampilkan melalui tabel laporan klasifikasi (*classification report*) yang merinci skor *Precision*, *Recall*, *F1-Score*, dan *accuracy*.

##### 4.5.1 Evaluasi Model Mobile JKN

Proses evaluasi model Mobile JKN dilakukan menggunakan tiga bagian dataset yaitu *training*, *validation*, dan *testing*. hasil evaluasi dapat dilihat sebagai berikut.

##### 1. Matrik evaluasi dataset *training* Mobile JKN

Tabel 4. 15 Matrik evaluasi dataset *training* Mobile JKN

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	0,99	1.00	0.99	708
<i>Usability</i>	<i>Negatif</i>	0.99	0.99	0.99	656
<i>Reliability</i>	<i>Positif</i>	1.00	1.00	1.00	664
<i>Reliability</i>	<i>Negatif</i>	0.99	0.99	0.99	704
<i>Efficiency</i>	<i>Positif</i>	1.00	1.00	1.00	702
<i>Efficiency</i>	<i>Negatif</i>	1.00	1.00	1.00	509
<i>Accuracy</i>				0.99	3943
<i>Macro Avg</i>		1.00	1.00	1.00	3943
<i>Weighted Avg</i>		0.99	0.99	0.99	3943

Pada hasil evaluasi dataset *training* tabel 4.15 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.99 (99%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif mencapai *Precision* 0.99, *Recall* 1.00, dan *F1-Score* 0.99. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 0.99.

b. *Reliability*

Pada aspek *reliability* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 0.99.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

2. Matrik evaluasi dataset *Validation* Mobile JKN

Tabel 4. 16 Matrik evaluasi dataset *validation* Mobile JKN

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	0.90	0.93	0.92	88
<i>Usability</i>	<i>Negatif</i>	0.84	0.90	0.87	81
<i>Reliability</i>	<i>Positif</i>	0.95	0.99	0.97	77
<i>Reliability</i>	<i>Negatif</i>	0.82	0.88	0.85	105
<i>Efficiency</i>	<i>Positif</i>	1.00	0.95	0.93	81
<i>Efficiency</i>	<i>Negatif</i>	0.91	1.00	0.98	69
<i>Accuracy</i>				0.96	501
<i>Macro Avg</i>		0.90	0.94	0.92	501
<i>Weighted Avg</i>		0.89	0.94	0.91	501

Pada hasil evaluasi dataset *validation* tabel 4.15 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.96 (96%), dengan nilai *Precision*,

*Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 0.90, *Recall* 0.93, dan *F1-Score* 0.92. Sedangkan untuk sentimen Negatif mencapai *Precision* 0.84, *Recall* 0.90, dan *F1-Score* 0.87.

b. *Reliability*

Pada aspek *reliability* sentimen Positif memperoleh *Precision* 0.95, *Recall* 0.99, dan *F1-Score* 0.97. Namun penurunan terlihat pada sentimen Negatif dengan *Precision* 0.82, *Recall* 0.88, dan *F1-Score* 0.85.

c. *Efficiency*

Pada aspek *efficiency* sentimen Positif memperoleh *Precision* 1.00, *Recall* 0.95, dan *F1-Score* 0.93. Sedangkan sentimen Negatif mencapai *Precision* 0.91, *Recall* 1.00, dan *F1-Score* 0.98.

3. Matrik evaluasi dataset *testing* Mobile JKN

Tabel 4. 17 Matrik evaluasi dataset *testing* Mobile JKN

<b>Aspek</b>	<b>Sentimen</b>	<b><i>Precision</i></b>	<b><i>Recall</i></b>	<b><i>F1-Score</i></b>	<b><i>Support</i></b>
<i>Usability</i>	<i>Positif</i>	0,93	0.99	0.96	83
<i>Usability</i>	<i>Negatif</i>	0.90	0.94	0.92	78
<i>Reliability</i>	<i>Positif</i>	0.93	0.98	0.95	86
<i>Reliability</i>	<i>Negatif</i>	0.75	0.87	0.80	85
<i>Efficiency</i>	<i>Positif</i>	0.97	0.98	0.97	87
<i>Efficiency</i>	<i>Negatif</i>	0.97	1.00	0.99	67
<i>Accuracy</i>				0.97	486
<i>Macro Avg</i>		0.91	0.96	0.93	486
<i>Weighted Avg</i>		0.91	0.96	0.93	486

Pada tabel evaluasi dataset *testing* 4.17 tahap pengujian akhir. Hasil ini merepresentasikan kemampuan generalisasi model yang menghasilkan *Accuracy* 0.97 (97%). dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* menunjukkan performa dengan sentimen Positif mencapai *Precision* 0.93, *Recall* 0.99, dan *F1-Score* 0.96. Untuk sentimen Negatif nilainya adalah *Precision* 0.90, *Recall* 0.94, dan *F1-Score* 0.92.

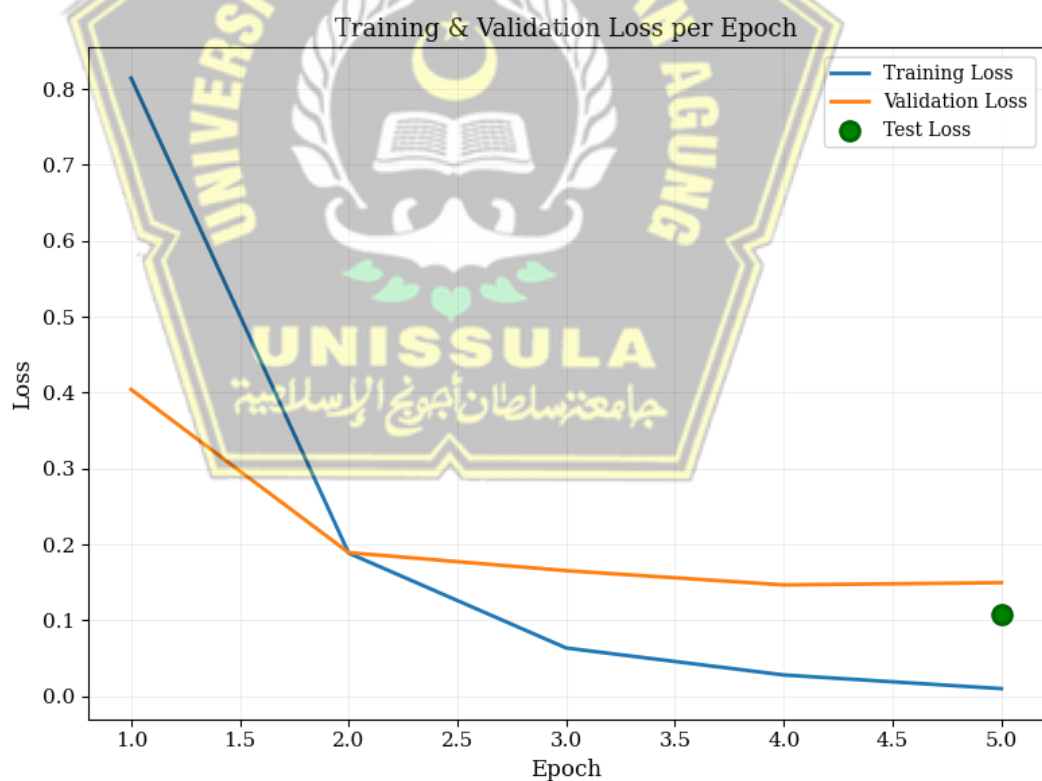
b. *Reliability*

Pada aspek *reliability* menunjukkan performa dengan sentimen Positif mencapai *Precision* 0.93, *Recall* 0.98, dan *F1-Score* 0.95. sedangkan sentimen Negatif memperoleh *Precision* terendah sebesar 0.75, *Recall* 0.87 dan *F1-Score* 0.80.

c. *Efficiency*

Pada aspek *efficiency* menunjukkan performa dengan sentimen Positif mencapai *Precision* 0.97, *Recall* 0.98, dan *F1-Score* 0.97. Sedangkan sentimen Negatif mencapai *Precision* 0.97, *Recall* 1.00, dan *F1-Score* 0.99.

4. Grafik *training dan validation loss* Mobile JKN



Gambar 4. 1 Grafik *Train Loss* dan *Validation Loss* Mobile JKN

Berdasarkan Gambar 4.1, visualisasi kurva *Loss* menunjukkan proses pembelajaran model selama 5 *epoch*. Training Loss mengalami penurunan

drastis dari angka di atas 0.8 pada *epoch* pertama menjadi mendekati 0.0 pada *epoch* kelima. Hal ini menunjukkan bahwa model sangat cepat beradaptasi dan mampu menyerap detail pola pada data latih dengan sangat baik.

Pada *validation loss* mengalami penurunan tajam hingga *epoch* ke-2, namun kemudian cenderung melandai di angka 0.15 hingga akhir proses. Terbentuknya jarak yang cukup lebar antara garis *training* dan *validation* setelah *epoch* ke-2 mengindikasikan adanya kecenderungan *overfitting* ringan, di mana model lebih bias pada data latih dibandingkan data validasi. Meskipun demikian, titik Test Loss yang rendah 0.1 mengonfirmasi bahwa model masih memiliki kemampuan generalisasi yang baik. Hal ini membuktikan bahwa model mempelajari data latih secara agresif, model tetap konsisten dan andal dalam memprediksi data baru yang belum pernah dilihat sebelumnya.

#### 5. Analisis Perbandingan Hasil Evaluasi Mobile JKN

Berdasarkan hasil dari tabel 4.15, 4.16, 4.17 *training*, *validation*, dan *testing* Aspek *efficiency* secara konsisten menjadi aspek yang paling mudah dikenali model baik difase validasi maupun *testing* dengan *F1-Score* diatas 0.97. Sebaliknya aspek *reliability* khususnya pada sentimen Negatif menjadi kelas yang paling sulit diklasifikasikan, terlihat dari nilai *Precision* di data *testing* yang menyentuh angka 0.75. Hal ini mengindikasikan adanya kemiripan fitur bahasa pada keluhan *reliability* dengan kelas lain, atau adanya ketidakseimbangan proporsi data yang mempengaruhi presisi prediksi model pada kelas tersebut dan ada 25% kemungkinan model salah tebak seperti tertukar dengan aspek lain atau sentimen positif. Pada gambar 4.1 grafik *Loss per Epoch* terlihat pola penurunan performa yang wajar dari fase *training* akurasi 0.99 ke fase *Testing* Akurasi 0.97. Selisih yang kecil ini menandakan model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting* yang parah, didukung oleh grafik *Validation Loss* yang tetap rendah dan tidak meningkat tajam di *epoch* akhir.

#### 4.5.2 Evaluasi Model M-Paspor

Proses evaluasi model M-Paspor dilakukan menggunakan tiga bagian dataset yaitu *training*, *validation*, dan *testing*. hasil evaluasi dapat dilihat sebagai berikut.

##### 1. Matriks Evaluasi data *training* M Paspor

Tabel 4. 18 Matrik evaluasi dataset *training* M Paspor

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	1.00	1.00	1.00	1363
<i>Usability</i>	<i>Negatif</i>	1.00	0.99	1.00	1113
<i>Reliability</i>	<i>Positif</i>	1.00	1.00	1.00	1254
<i>Reliability</i>	<i>Negatif</i>	1.00	1.00	1.00	1539
<i>Efficiency</i>	<i>Positif</i>	1.00	1.00	1.00	1260
<i>Efficiency</i>	<i>Negatif</i>	1.00	1.00	1.00	886
<i>Accuracy</i>				0.96	7415
<i>Macro Avg</i>		1.00	1.00	1.00	7415
<i>Weighted Avg</i>		1.00	1.00	1.00	7415

Pada hasil evaluasi dataset *training* tabel 4.18 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.96 (96%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

##### a. *Usability*

Pada aspek *usability* sentimen Positif mencapai nilai sempurna dengan *Precision* 1.00, *Recall* 1.00, dan *F1-Score* 1.00. Sedangkan sentimen Negatif memperoleh *Precision* 1.00, *Recall* 0.99, dan *F1-Score* 1.00.

##### b. *Reliability*

Pada aspek *reliability* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

##### c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

## 2. Matriks Evaluasi data *Validation* M Paspor

Tabel 4. 19 Matrik evaluasi data *validation* M Paspor

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	1.00	0.99	0.99	177
<i>Usability</i>	<i>Negatif</i>	0.90	0.96	0.93	127
<i>Reliability</i>	<i>Positif</i>	1.00	1.00	1.00	174
<i>Reliability</i>	<i>Negatif</i>	0.88	0.93	0.91	181
<i>Efficiency</i>	<i>Positif</i>	1.00	1.00	1.00	157
<i>Efficiency</i>	<i>Negatif</i>	1.00	1.00	1.00	93
<i>Accuracy</i>				0.98	909
<i>Macro Avg</i>		0.96	0.98	0.97	909
<i>Weighted Avg</i>		0.96	0.98	0.97	909

Pada hasil evaluasi dataset *validation* tabel 4.19 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.98 (98%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

### a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 1.00, *Recall* 0.99, dan *F1-Score* 0.99. Untuk sentimen Negatif terjadi sedikit variasi namun tetap baik dengan *Precision* 0.90, *Recall* 0.96, dan *F1-Score* 0.93.

### b. *Reliability*

Pada aspek *reliability* sentimen Positif sangat stabil dengan nilai sempurna 1.00 untuk semua metrik. Sedangkan sentimen Negatif memperoleh *Precision* sebesar 0.88, *Recall* 0.93, dan *F1-Score* 0.91.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

3. Matriks Evaluasi data *testing* M Paspur

Tabel 4. 20 Matriks Evaluasi data *testing* M Paspur

<b>Aspek</b>	<b>Sentimen</b>	<b><i>Precision</i></b>	<b><i>Recall</i></b>	<b><i>F1-Score</i></b>	<b><i>Support</i></b>
<i>Usability</i>	<i>Positif</i>	1.00	0.99	1.00	168
<i>Usability</i>	<i>Negatif</i>	0.92	0.96	0.94	118
<i>Reliability</i>	<i>Positif</i>	1.00	1.00	1.00	166
<i>Reliability</i>	<i>Negatif</i>	0.83	0.93	0.91	180
<i>Efficiency</i>	<i>Positif</i>	0.99	0.99	0.99	181
<i>Efficiency</i>	<i>Negatif</i>	0.99	1.00	0.99	83
<i>Accuracy</i>				0.98	896
<i>Macro Avg</i>		0.96	0.98	0.97	896
<i>Weighted Avg</i>		0.96	0.98	0.97	896

Pada hasil evaluasi dataset *testing* tabel 4.20 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.98 (98%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 1.00, *Recall* 0.99, dan *F1-Score* 1.00. Untuk sentimen Negatif terjadi sedikit variasi namun tetap baik dengan *Precision* 0.92, *Recall* 0.96, dan *F1-Score* 0.94.

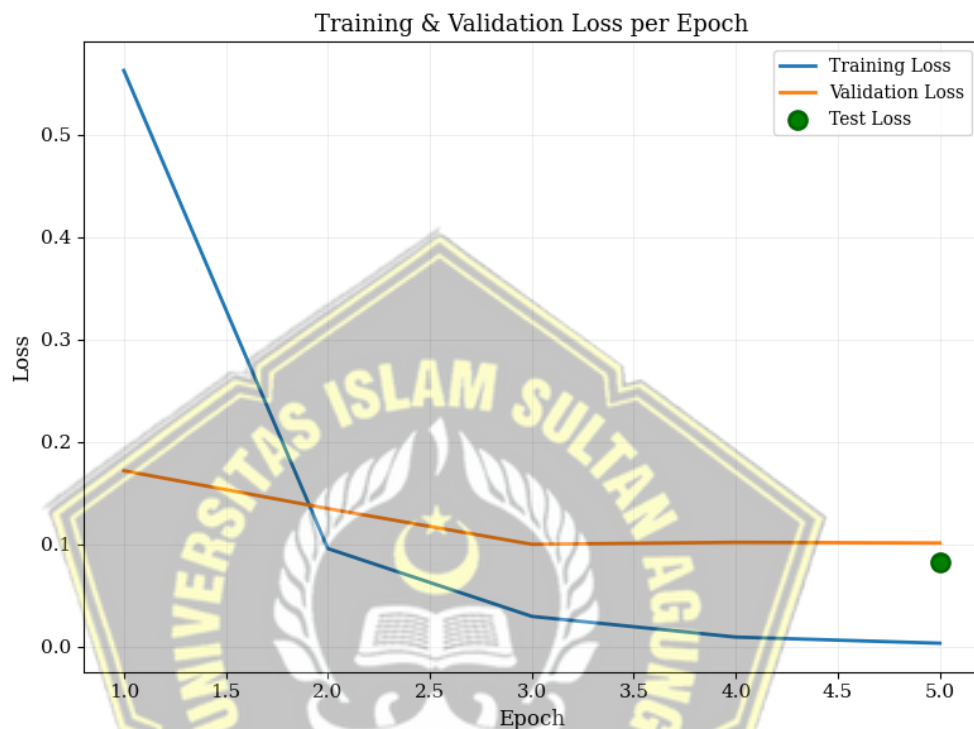
b. *Reliability*

Pada aspek *reliability* sentimen Positif sangat stabil dengan nilai sempurna 1.00 untuk semua metrik. Sedangkan sentimen Negatif memperoleh *Precision* 0.83, *Recall* 0.93, dan *F1-Score* 0.91.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision* 0.99, *Recall* 1.00, dan *F1-Score* 0.99.

#### 4. Grafik *training* dan *validation loss* M Paspor



Gambar 4. 2 Grafik *training* dan *validation loss* M Paspor

Berdasarkan Gambar 4.2 visualisasi kurva *Loss* menunjukkan proses pembelajaran model selama 5 *epoch*. Dari angka 0.55 pada *epoch* pertama hingga menyentuh angka mendekati 0.0 pada *epoch* kelima. Hal ini menunjukkan bahwa model sangat cepat beradaptasi dan mengenali pola pada data latih. *Validation Loss* dimulai dari angka yang sudah cukup rendah sekitar 0.17 dan terus menurun secara perlahan hingga stabil di kisaran 0.10. Titik *Test Loss* yang berada di posisi rendah 0.08 mengonfirmasi bahwa performa model pada data uji selaras dengan data validasi, membuktikan konsistensi model dalam menangani data baru.

#### 5. Analisis Perbandingan Hasil Evaluasi M Paspor

Berdasarkan hasil dari tabel 4.18, 4.19, 4.20 aspek *Efficiency* adalah aspek yang paling mudah dipelajari oleh model. Hal ini dibuktikan dengan nilai F1-

Score yang konsisten di angka 0.99 hingga 1.00 baik di fase validasi maupun *testing*. Ini mengindikasikan bahwa ulasan terkait efisiensi pada M-Paspor memiliki pola kata yang sangat khas dan tidak ambigu. Pada gambar 4.2 grafik *Loss per Epoch* model menunjukkan stabilitas yang luar biasa. Tidak terjadi penurunan performa yang signifikan dari fase *training* ke *validation* maupun *testing*. Akurasi yang bertahan di angka 0.98 pada data *testing* membuktikan model tidak mengalami *Overfitting*. Meskipun secara umum sangat baik, model memiliki sedikit kesulitan pada sentimen Negatif di aspek *Reliability*, terlihat dari nilai *Precision testing* sebesar 0.83 paling rendah dibanding kelas lain. Hal ini disebabkan oleh irisan makna pada keluhan teknis aplikasi M-Paspor yang terkadang mirip dengan keluhan *usability*, sehingga model sedikit bias dalam memisahkan kedua kelas tersebut.

#### 4.5.3 Evaluasi Model SIGNAL

Proses evaluasi model SIGNAL dilakukan menggunakan tiga bagian dataset yaitu *training*, *validation*, dan *testing*. hasil evaluasi dapat dilihat sebagai berikut.

##### 1. Matriks Evaluasi data *training* SIGNAL

Tabel 4. 21 Matrik evaluasi dataset *training* SIGNAL

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	0.99	1.00	1.00	1403
<i>Usability</i>	<i>Negatif</i>	1.00	1.00	1.00	927
<i>Reliability</i>	<i>Positif</i>	1.00	1.00	1.00	1007
<i>Reliability</i>	<i>Negatif</i>	1.00	1.00	1.00	914
<i>Efficiency</i>	<i>Positif</i>	1.00	1.00	1.00	1085
<i>Efficiency</i>	<i>Negatif</i>	1.00	1.00	1.00	711
<i>Accuracy</i>				0.99	6047
<i>Macro Avg</i>		1.00	1.00	1.00	6047
<i>Weighted Avg</i>		1.00	1.00	1.00	6047

Pada hasil evaluasi dataset *training* tabel 4.21 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.99 (99%), dengan nilai *Precision*,

*Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 0.99, *Recall* 1.00, dan *F1-Score* 1.00. Sedangkan Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

b. *Reliability*

Pada aspek *reliability* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif untuk *Precision*, *Recall*, dan *F1-Score* mencapai 1.00 secara konsisten. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

2. Matriks Evaluasi data *validation* SIGNAL

Tabel 4. 22 Matrik Evaluasi data *validation* SIGNAL

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	0.94	0.95	0.94	177
<i>Usability</i>	<i>Negatif</i>	1.00	1.00	1.00	123
<i>Reliability</i>	<i>Positif</i>	0.92	1.00	0.96	117
<i>Reliability</i>	<i>Negatif</i>	0.92	0.93	0.92	99
<i>Efficiency</i>	<i>Positif</i>	0.93	0.95	0.94	151
<i>Efficiency</i>	<i>Negatif</i>	0.99	1.00	0.99	97
<i>Accuracy</i>				0,98	764
<i>Macro Avg</i>		0.95	0.97	0.96	764
<i>Weighted Avg</i>		0.95	0.97	0.96	764

Pada hasil evaluasi dataset *validation* tabel 4.21 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.98 (98%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 0.94, *Recall* 0.95, dan *F1-Score* 0.94. Sedangkan Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

b. *Reliability*

Pada aspek *reliability* Sentimen Positif memperoleh *Precision* 0.92, *Recall* 1.00, dan *F1-Score* 0.96. Untuk sentimen Negatif memperoleh *Precision* 0.92, *Recall* 0.93, dan *F1-Score* 0.92.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif meraih *Precision* 0.93, *Recall* 0.95, dan *F1-Score* 0.94. Untuk sentimen memperoleh *Precision* 0.99, *Recall* 1.00, dan *F1-Score* 0.99.

3. Matriks Evaluasi data *testing* SIGNAL

Tabel 4. 23 Matrik Evaluasi data *testing* SIGNAL

Aspek	Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Usability</i>	<i>Positif</i>	0.92	0.94	0.93	188
<i>Usability</i>	<i>Negatif</i>	0.97	1.00	0.99	103
<i>Reliability</i>	<i>Positif</i>	0.89	0.98	0.93	124
<i>Reliability</i>	<i>Negatif</i>	0.96	0.98	0.97	121
<i>Efficiency</i>	<i>Positif</i>	0.93	0.95	0.94	147
<i>Efficiency</i>	<i>Negatif</i>	1.00	1.00	1.00	78
<i>Accuracy</i>				0,98	761
<i>Macro Avg</i>		0.95	0.97	0.96	761
<i>Weighted Avg</i>		0.94	0.97	0.95	761

Pada hasil evaluasi dataset *testing* tabel 4.23 menghasilkan performa pada dengan nilai *accuracy* mencapai 0.98 (98%), dengan nilai *Precision*, *Recall*, dan *F1-Score* pada aspek *Usability*, *Reliability*, *Efficiency* sebagai berikut.

a. *Usability*

Pada aspek *usability* sentimen Positif memperoleh *Precision* 0.92, *Recall* 0.94, dan *F1-Score* 0.93. Sedangkan untuk sentimen Negatif memperoleh *Precision* 0.97, *Recall* 1.00, dan *F1-Score* 0.99.

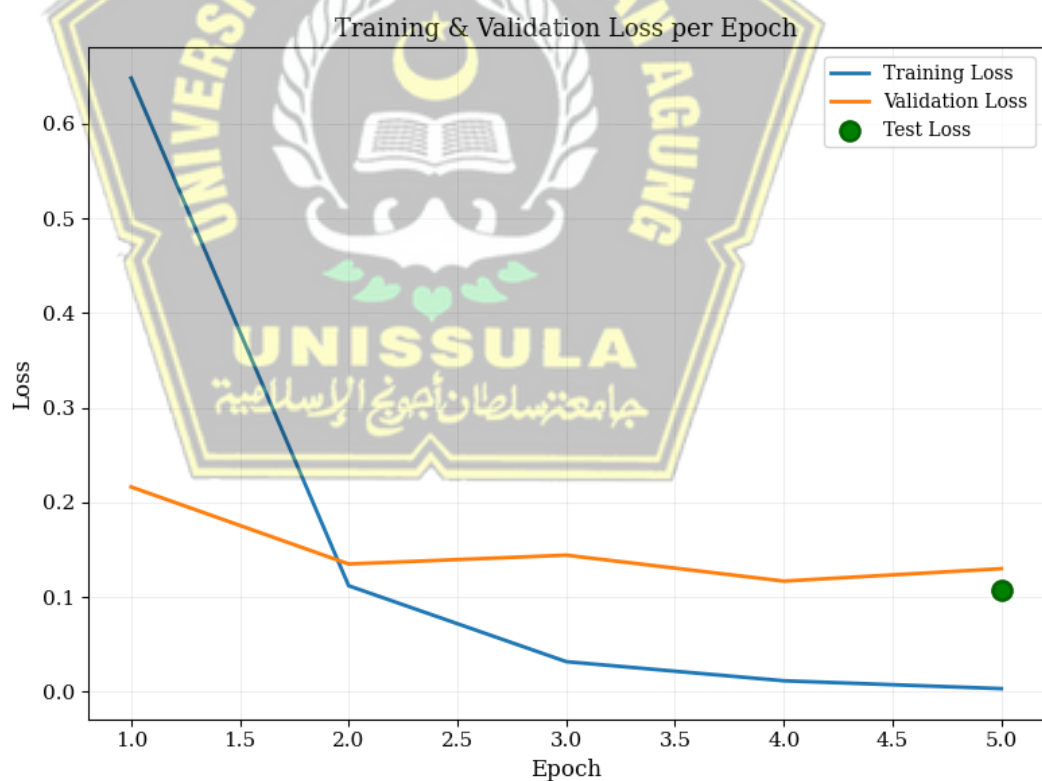
b. *Reliability*

Pada aspek *reliability* sentiment Positif memperoleh *Precision* 0.89, *Recall* 0.98 dan *F1-Score* 0.93. Untuk sentimen Negatif dengan memperoleh *Precision* 0.96, *Recall* 0.98, dan *F1-Score* 0.97.

c. *Efficiency*

Pada aspek *Efficiency* sentimen Positif meraih *Precision* 0.93, *Recall* 0.95, dan *F1-Score* 0.94. Untuk sentimen Negatif nilai *Precision*, *Recall*, dan *F1-Score* konsisten diangka 1.00.

4. Grafik *training* dan *validation loss* SIGNAL



Gambar 4. 3 Grafik *training* dan *validation loss* SIGNAL

Berdasarkan Gambar 4.2 visualisasi kurva *Loss* menunjukkan proses pembelajaran model selama 5 *epoch*. Dimulai dari angka sekitar 0.65 pada

*epoch* pertama dan turun drastis hingga mendekati 0.0 pada *epoch* kelima. penurunan ini mengindikasikan bahwa model dapat menyerap informasi dan pola dari data latih SIGNAL dengan sangat cepat. *Validation Loss* dimulai dari posisi yang cukup rendah yaitu 0.22, kemudian menurun perlahan dan stabil di kisaran angka 0.12. Grafik ini menunjukkan tren yang sehat karena garis oranye terus melandai dan tidak menjauh dari garis biru, menandakan bahwa model tidak mengalami *overfitting*. Titik *Test Loss* yang berada di posisi rendah 0.10. Posisi titik ini sedikit lebih rendah dibandingkan akhir garis validasi mengonfirmasi bahwa performa model model justru bekerja lebih baik atau setidaknya sama baiknya saat menghadapi data *testing* dibandingkan saat fase validasi.

#### 5. Analisis Perbandingan Hasil Evaluasi SIGNAL

Berdasarkan hasil dari tabel 4.21, 4.22, 4.23 model justru sangat unggul dalam mendeteksi sentimen negatif. Hal ini terlihat dari nilai *F1-Score* memperoleh 1.00 pada aspek *Efficiency* Negatif dan skor 0.99 pada *Usability* Negatif saat *testing*. Hal ini mengindikasikan ulasan negatif pada aplikasi SIGNAL memiliki pola kalimat yang sangat tegas dan eksplisit, sehingga sangat mudah dikenali oleh model. Terdapat sedikit bias pada aspek *Reliability* sentimen Positif yang mencatatkan nilai *Precision* sebesar 0.89. Ini mengindikasikan bahwa ada ambiguitas di mana model memprediksi ulasan sebagai *Reliability* Positif padahal sebenarnya masuk kategori lain.

#### 4.6 Analisis Performa Model

Setelah dilakukan evaluasi model pada masing-masing dataset Mobile JKN, M-Paspor dan SIGNAL, tahap selanjutnya adalah melakukan analisis perbandingan performa model secara komprehensif. Analisis ini bertujuan untuk mengidentifikasi kekuatan dan kelemahan model dalam mengklasifikasikan sentimen dan aspek pada setiap dataset, serta memahami karakteristik unik dari masing-masing aplikasi yang mempengaruhi performa klasifikasi.

Perbandingan dilakukan dengan menganalisis berbagai metrik evaluasi yang meliputi *accuracy*, *precision*, *recall*, dan *F1-Score* pada fase *testing* dari ketiga dataset. Selain itu, dilakukan juga analisis terhadap *loss function* untuk memahami proses pembelajaran model dan kemampuan generalisasinya. Perbandingan ini penting untuk mengevaluasi konsistensi performa model IndoBERT dalam menangani data ulasan aplikasi yang berbeda-beda karakteristiknya, serta memberikan rekomendasi untuk implementasi praktis di lapangan.

Analisis perbandingan disajikan dalam beberapa tabel yang mencakup perbandingan akurasi keseluruhan, performa per aspek dan sentimen, *precision* terendah sebagai indikator kesulitan klasifikasi, karakteristik pembelajaran model berdasarkan *loss function*, serta ringkasan metrik *macro* dan *weighted average*. Setiap tabel dilengkapi dengan penjelasan dan interpretasi untuk memberikan pemahaman mendalam terkait pola yang ditemukan.

1. Perbandingan *accuracy* model Mobile JKN, M-Paspor, SIGNAL

Tabel 4. 24 Perbandingan *accuracy* model Mobile JKN, M-Paspor, SIGNAL

<b>Dataset</b>	<b>Training Accuracy</b>	<b>Validation Accuracy</b>	<b>Testing Accuracy</b>
Mobile JKN	0.99 (99%)	0.96 (96%)	0.97 (97%)
M-Paspor	0.96 (96%)	0.98 (98%)	0.98 (98%)
SIGNAL	0.99 (99%)	0.98 (98%)	0.98 (98%)

Berdasarkan Tabel 4.24 dataset M-Paspor dan SIGNAL menunjukkan akurasi *testing* sebesar 0.98, sementara Mobile JKN sedikit lebih rendah pada 0.97. M-Paspor menunjukkan pola peningkatan dari *training* 0.96 ke *validation* dan *testing* 0.98, mengindikasikan model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik. Mobile JKN mengalami penurunan wajar dari *training* 0.99 ke *validation* 0.96, namun kembali meningkat ke *testing* 0.97, menunjukkan stabilitas model. SIGNAL mempertahankan konsistensi dari *validation* hingga *testing* pada angka 0.98.

## 2. Perbandingan *F1-Score* per aspek pada Data *Testing*

Tabel 4. 25 Perbandingan *F1-Score* per aspek pada Data *Testing*

Aspek	Sentimen	Mobile JKN	M-Paspor	SIGNAL
<i>Usability</i>	Positif	0.96	1.00	0.93
<i>Usability</i>	Negatif	0.92	0.94	0.99
<i>Reliability</i>	Positif	0.95	1.00	0.93
<i>Reliability</i>	Negatif	0.80	0.91	0.97
<i>Efficiency</i>	Positif	0.97	0.99	0.94
<i>Efficiency</i>	Negatif	0.99	0.99	1.00

Berdasarkan Tabel 4.25 aspek *Efficiency* menunjukkan performa tertinggi di semua dataset dengan *F1-Score* di atas 0.94, bahkan SIGNAL mencapai nilai sempurna 1.00 untuk sentimen Negatif. Hal ini mengonfirmasi bahwa ulasan terkait efisiensi memiliki pola kalimat yang khas dan mudah dibedakan. Aspek *Reliability* menunjukkan variasi performa yang signifikan. Mobile JKN memiliki *F1-Score* terendah pada sentimen Negatif (0.80), M-Paspor di posisi tengah (0.91), dan SIGNAL tertinggi (0.97). Untuk aspek *Usability*, M-Paspor unggul dalam sentimen Positif (1.00), sementara SIGNAL unggul dalam sentimen Negatif (0.99).

## 3. Perbandingan *Precision* terendah model Mobile JKN, M Paspor, Signal pada dataset *testing*

Tabel 4. 26 Perbandingan *Precision*

Dataset	Aspek - Sentimen	<i>Precision</i>
Mobile JKN	<i>Reliability</i> - Negatif	0.75
M-Paspor	<i>Reliability</i> - Negatif	0.83
SIGNAL	<i>Reliability</i> - Positif	0.89

Berdasarkan Tabel 4.26 Mobile JKN memiliki nilai *precision* terendah sebesar 0,75 pada aspek *Reliability* dengan sentimen negatif. Hal ini menandakan adanya kata-kata yang digunakan seringkali mirip dengan kelas lain. Analisis sebelumnya menunjukkan bahwa keluhan terkait *reliability* di

Mobile JKN sering mirip dengan keluhan mengenai *usability* atau aspek lainnya. Sementara itu, M-Paspor mencatat *precision* 0,83 menunjukkan adanya irisan makna antara keluhan teknis mengenai *reliability* dan masalah *usability*, seringkali keluhan masuk kategori *reliability* juga berkaitan dengan *usability*. Sedangkan SIGNAL menampilkan pola berbeda, dengan kesulitan terlihat pada *Reliability* positif yang mencapai 0,89 hal ini karena ungkapan kepuasan cenderung lebih umum dan kurang spesifik.

4. Perbandingan *Test Loss* dan Karakteristik Model Mobile JKN, M-Paspor, Signal

Tabel 4. 27 Perbandingan *Test Loss* dan Karakteristik Model

<b>Dataset</b>	<b>Test Loss</b>	<b>Training Loss</b>	<b>Validation Loss</b>
Mobile JKN	0.10	0.8 - 0.0	0.4 - 0.15
M-Paspor	0.08	0.55 - 0.0	0.17 - 0.10
SIGNAL	0.10	0.65 - 0.0	0.22 - 0.12

Berdasarkan Tabel 4.27 M-Paspor memiliki *test loss* terendah 0.08, mengonfirmasi performa mendapatkan tingkat error terendah. Mobile JKN dan SIGNAL dengan *test loss* 0.10 menunjukkan performa yang sangat baik dan setara. Ketiga dataset menunjukkan pola penurunan *training loss* yang drastis dalam 5 *epoch*, mengindikasikan model IndoBERT dapat beradaptasi dengan cepat. *Validation loss* yang stabil dan tidak meningkat di *epoch* akhir membuktikan model tidak mengalami *overfitting* yang signifikan.

5. Perbandingan *Macro Average* dan *Weighted Average dataset Testing*

Tabel 4. 28 Perbandingan *Macro Average* dan *Weighted Average dataset Testing*

<b>Dataset</b>	<b>Macro Avg Precision</b>	<b>Macro Avg Recall</b>	<b>Macro Avg F1-Score</b>	<b>Weighted Avg F1-Score</b>
Mobile JKN	0.91	0.96	0.93	0.93
M-Paspor	0.96	0.98	0.97	0.97
SIGNAL	0.95	0.97	0.96	0.95

Berdasarkan Tabel 4.28 M-Paspor memperoleh metrik dengan *Macro Avg F1-Score* 0.97 dan *Weighted Avg* 0.97, menunjukkan performa seimbang di semua kelas. SIGNAL berada di posisi kedua dengan *Macro Avg* 0.96, sementara

Mobile JKN di posisi ketiga dengan 0.93. Perbedaan minimal antara *Macro* dan *Weighted Average* di semua dataset menunjukkan distribusi kelas yang relatif seimbang dan tidak ada bias signifikan terhadap kelas mayoritas

#### 6. Analisis performa Keseluruhan Antar Dataset

Dataset M-Paspor menunjukkan performa terbaik dengan *testing accuracy* 0.98, *test loss* terendah 0.08, dan *macro avg F1-Score* tertinggi 0.97. Keunggulan ini disebabkan data M-Paspor memiliki ulasan yang terstruktur dan spesifik terkait proses pengajuan paspor, sehingga model mudah mengenali pola yang konsisten. SIGNAL mempertahankan akurasi 0.98 karena pengguna mengekspresikan keluhan secara eksplisit dan tegas. Mobile JKN mencapai *testing accuracy* 0.97 sedikit lebih rendah karena variasi topik layanan kesehatan yang lebih beragam menyebabkan keberagaman dalam ulasan.

Model menunjukkan kemampuan generalisasi yang baik pada ketiga dataset dibuktikan dengan selisih kecil antara akurasi *training* dan *testing*. Mobile JKN memiliki selisih terbesar yaitu 2% dari 0.99 menjadi 0.97. M-Paspor mengalami peningkatan akurasi dari *training* 0.96 ke *testing* 0.98, menunjukkan model justru lebih baik menggeneralisasi data baru. SIGNAL mempertahankan konsistensi tertinggi dengan akurasi stabil di 0.98 dari *validation* hingga *testing*. *Validation loss* yang menurun stabil tanpa peningkatan tajam di *epoch* akhir pada semua dataset membuktikan model tidak mengalami *overfitting* dan *hyperparameter* yang digunakan sudah optimal.

Aspek *Efficiency* konsisten menunjukkan performa tertinggi *F1-Score* 0.94 pada semua dataset karena pengguna menggunakan kata-kata spesifik seperti "lemot", "cepat", "lambat" yang memiliki makna semantik jelas dan tidak tumpang tindih dengan aspek lainnya. Aspek *Usability* bervariasi antar dataset M-Paspor mencapai 1.00 karena alur aplikasi yang linear, sementara SIGNAL hanya 0.93 karena berbagai fitur kompleks menyebabkan *feedback* lebih beragam dan overlap dengan *reliability*.

Aspek *Reliability* menjadi tantangan utama dengan pola berbeda di setiap dataset. Mobile JKN kesulitan pada sentimen Negatif *precision* 0.75 karena keluhan *reliability* sering diekspresikan bersama keluhan *usability* contoh

"aplikasi error terus, susah masuk". M-Paspor lebih baik *precision* 0.83 karena keluhan lebih spesifik seperti "gagal upload" atau "pembayaran error". SIGNAL berbeda kesulitan pada sentimen Positif *precision* 0.89 karena pujian yang general seperti "aplikasi bagus" tanpa menyebutkan aspek spesifik, sementara keluhan negatif sangat eksplisit sehingga mudah diklasifikasi.

#### **4.7 Hasil Implementasi Antarmuka**

Setelah tahap pemodelan selesai dan model telah menunjukkan performa yang optimal, tahap berikutnya dalam penelitian ini adalah implementasi sistem ke dalam sebuah platform yang dapat diakses oleh pengguna. Untuk tujuan ini, dipilih *framework Streamlit* yang memungkinkan pembangunan antarmuka web yang interaktif dan dinamis dengan menggunakan bahasa pemrograman *Python*. Streamlit dipilih karena kemudahannya dalam integrasi dengan model *machine learning*.

##### **4.7.1 Hasil Perancangan Antarmuka**

Analisis teks memungkinkan pengguna untuk menganalisis sentiment berbasis aspek dari teks yang dimasukkan secara langsung. Terdapat sebuah kotak input tempat pengguna dapat mengetik atau menempelkan teks yang ingin dianalisis. Setelah teks dimasukkan, pengguna dapat memilih aplikasi pelayanan masyarakat terlebih dahulu dengan opsi mobile jkn, mpaspor, dan signal untuk memproses teks dan mendapatkan hasil analisis sentimen, untuk mengkategorikan sentiment dalam teks menjadi berdasarkan aspek dan sentiment yang telah ditentukan. Pada gambar 4.4 merupakan tampilan awal analisis teks

## ID Aspect-Based Sentiment Analysis (ABSA) Pada Aplikasi Pelayanan Masyarakat

Masukkan kalimat ulasan, pilih model, dan lihat aspek serta sentimennya.

Pilih model ABSA yang digunakan:

MPaspor

Masukkan kalimat ulasan Anda:

Gambar 4. 4 Tampilan *Interface*

Pada tahap ini pengguna dapat memilih salah satu aplikasi pelayanan masyarakat yang telah disediakan dalam sistem, misalnya Mobile JKN, M-Paspor, atau SIGNAL. Setelah aplikasi dipilih, pengguna dapat memasukkan teks ulasan atau opini terkait aplikasi tersebut pada kolom input yang tersedia. Sistem kemudian akan melakukan proses analisis dengan mendeteksi aspek yang terkandung dalam teks, sekaligus menentukan sentimen yang menyertainya.

Pilih model ABSA yang digunakan:

JKN

Masukkan kalimat ulasan Anda:

aplikasi susah digunakan dan verifikasi wajah gagal terus

✓ Aspek terdeteksi dari model JKN:

- Usability (negatif): "aplikasi susah digunakan"
- Reliability (negatif): "verifikasi wajah gagal terus"

Gambar 4. 5 Tampilan hasil deteksi multiaspek

Berdasarkan Gambar 4.5 pengujian sistem dengan ulasan "aplikasi susah digunakan dan verifikasi wajah gagal terus" menunjukkan keberhasilan

implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi multi-aspek dari satu kalimat ulasan untuk aplikasi Mobile JKN. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Usability* (negatif) terdeteksi dari frasa "aplikasi susah digunakan" dan *Reliability* (negatif) terdeteksi dari frasa "verifikasi wajah gagal terus".



Pilih model ABSA yang digunakan:

JKN

Masukkan kalimat ulasan Anda:

aplikasi susah digunakan

✓ Aspek terdeteksi dari model JKN:

- *Usability* (negatif): "aplikasi susah digunakan"

Gambar 4. 6 Tampilan hasil deteksi *single* aspek pada model JKN

Berdasarkan Gambar 4.6 pengujian sistem dengan ulasan "aplikasi susah digunakan" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi single-aspek dari satu kalimat ulasan untuk aplikasi Mobile JKN. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Usability* (negatif) terdeteksi dari frasa "aplikasi susah digunakan".

Pilih model ABSA yang digunakan:

MPaspor

Masukkan kalimat ulasan Anda:

pembuatan paspor lama dan aplikasi tidak bisa dibuka

Aspek terdeteksi dari model MPaspor:

- **Efficiency (negatif):** "pembuatan paspor lama"
- **Reliability (negatif):** "dan aplikasi tidak bisa dibuka"

Gambar 4. 7 Tampilan hasil deteksi multiaspek pada model M-Paspor

Berdasarkan Gambar 4.7 pengujian sistem dengan ulasan "pembuatan paspor lama dan aplikasi tidak bisa dibuka" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi multi-aspek dari satu kalimat ulasan untuk aplikasi M-Paspor. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Effeciency* (negatif) terdeteksi dari frasa "pembuatan passport lama" dan *Reliability* (negatif) terdeteksi dari frasa "aplikasi tidak bisa dibuka".

Pilih model ABSA yang digunakan:

MPaspor

Masukkan kalimat ulasan Anda:

Aplikasi suka keluar sendiri

Aspek terdeteksi dari model MPaspor:

- **Reliability (negatif):** "Aplikasi suka keluar sendiri"

Gambar 4. 8 Tampilan hasil deteksi single aspek pada model M-Paspor

Berdasarkan Gambar 4.7 pengujian sistem dengan ulasan "aplikasi suka keluar sendiri" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi single-aspek dari satu kalimat ulasan untuk aplikasi M-Paspor. Hasil deteksi menunjukkan bahwa

model berhasil mengidentifikasi single aspek beserta sentimennya secara akurat yaitu *Reliability* (negatif) terdeteksi dari frasa "aplikasi tidak bisa dibuka".

Pilih model ABSA yang digunakan:

Signal

Masukkan kalimat ulasan Anda:

tampilan aplikasi bagus tapi lumayan lama saat dibuka

✓ Aspek terdeteksi dari model Signal:

- Usability (positif): "tampilan aplikasi bagus"
- Efficiency (negatif): "lumayan lama saat dibuka"

Gambar 4. 9 Tampilan hasil deteksi multi aspek pada model SIGNAL

Berdasarkan Gambar 4.9 pengujian sistem dengan ulasan "tampilan aplikasi bagus tapi lumayan lama saat dibuka" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi multi-aspek dari satu kalimat ulasan untuk aplikasi SIGNAL. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Usability* (negatif) terdeteksi dari frasa "tampilan aplikasi bagus" dan *efficiency* (negatif) terdeteksi dari frasa "lumayan lama saat dibuka".

Pilih model ABSA yang digunakan:

Signal

Masukkan kalimat ulasan Anda:

tampilan aplikasi jelek

✓ Aspek terdeteksi dari model Signal:

- Usability (negatif): "tampilan aplikasi jelek"

Gambar 4. 10 Tampilan hasil deteksi single aspek pada model SIGNAL

Berdasarkan Gambar 4.10 pengujian sistem dengan ulasan "tampilan aplikasi jelek" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan Streamlit ini mampu melakukan deteksi single-aspek dari

satu kalimat ulasan untuk aplikasi SIGNAL. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Usability* (negatif) terdeteksi dari frasa "tampilan aplikasi jelek".

#### 4.8 Pembahasan

Penelitian ini berhasil mengembangkan model *Aspect-Based Sentiment Analysis* (ABSA) menggunakan arsitektur IndoBERT pada tiga dataset aplikasi pelayanan publik Mobile JKN, M-Paspor, dan SIGNAL. Model menunjukkan kinerja yang sangat baik dengan *testing accuracy* mencapai 97% untuk Mobile JKN, 98% untuk M-Paspor, dan 98% untuk SIGNAL. Tingginya performa ini menunjukkan bahwa pendekatan IndoBERT yang menggabungkan pemahaman konteks bahasa Indonesia dengan teknik *token classification* sangat efektif untuk tugas ABSA pada domain aplikasi pelayanan publik. Kemampuan generalisasi model terbukti baik dengan selisih minimal antara *training* dan *testing accuracy*, di mana Mobile JKN mengalami penurunan 2% dari 0.99 ke 0.97, M-Paspor justru meningkat dari 0.96 ke 0.98, dan SIGNAL stabil di 0.98. *Validation loss* yang menurun stabil tanpa peningkatan tajam di *epoch* akhir mengonfirmasi model tidak mengalami *overfitting*, menunjukkan *hyperparameter* yang digunakan sudah optimal untuk ketiga dataset.

Secara keseluruhan temuan ini menunjukkan bahwa meskipun aplikasi pelayanan publik telah berupaya meningkatkan kemudahan penggunaan (*Usability*) dan aspek keandalan (*Reliability*) secara konsisten menjadi titik lemah yang paling banyak dikeluhkan pengguna di ketiga platform. Selain evaluasi model implementasi berbasis Streamlit dibuat untuk memudahkan pengguna dalam melakukan analisis sentiment berbasis aspek. Sistem ini memiliki fitur utama yaitu analisis teks yang memungkinkan pengguna memasukkan teks secara langsung untuk dianalisis sentimennya. Implementasi ini memberikan kemudahan dalam memahami dan mengakses hasil analisis secara *real time*, baik pada ulasan dengan satu aspek (*single-aspect*) maupun ulasan kompleks yang mengandung banyak aspek (*multi-aspect*).

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian mengenai analisis sentimen berbasis aspek pada aplikasi pelayanan masyarakat menggunakan IndoBERT telah menghasilkan beberapa kesimpulan berikut:

1. *Aspect-Based Sentiment Analysis* (ABSA) telah berhasil dirancang dan diimplementasikan dengan mengkombinasikan arsitektur model IndoBERT dan pendekatan token classification. Alur sistem yang mencakup pengumpulan data, pra-pemrosesan, pelabelan, penyeimbangan data, hingga *fine-tuning* model terbukti mampu mengekstraksi aspek dan sentimen dari ulasan pengguna secara simultan dan efektif.
2. Performa model IndoBERT yang dikembangkan untuk tugas ABSA menunjukkan kinerja yang sangat baik dan solid. Hal ini dibuktikan dengan pencapaian *testing accuracy* yang tinggi pada ketiga aplikasi, yaitu 97% untuk Mobile JKN, 98% untuk M-Paspor, dan 98% untuk SIGNAL. Kinerja yang kuat ini juga tercermin pada metrik *precision*, *recall*, dan *F1-score* untuk setiap aspek dan sentimen, yang menandakan kemampuan model dalam mengklasifikasikan setiap kelas aspek sentimen dengan andal. Model menunjukkan kemampuan generalisasi yang baik tanpa *overfitting*, dibuktikan dengan selisih minimal antara training dan *testing accuracy* serta *validation loss* yang stabil.
3. Analisis perbandingan performa model antar dataset mengungkapkan karakteristik unik masing-masing aplikasi. Aspek *Efficiency* secara konsisten menjadi yang paling mudah diklasifikasi di semua dataset dengan *F1-Score* di atas 0.94, sementara aspek *Reliability* menjadi tantangan utama dengan pola kesulitan berbeda Mobile JKN kesulitan pada sentimen Negatif dengan *precision* 0.75, M-Paspor pada sentimen Negatif *precision* 0.83, dan SIGNAL pada sentimen Positif *precision* 0.89. Perbedaan ini disebabkan oleh

kompleksitas domain, eksplisitas ekspresi pengguna, dan variasi layanan pada masing-masing aplikasi.

4. Implementasi sistem berbasis *Streamlit* berhasil mengoperasionalkan model ABSA untuk penggunaan praktis, memungkinkan analisis sentimen berbasis aspek secara *real-time* dengan antarmuka yang *user-friendly*. Sistem mampu menangani ulasan *single-aspect* maupun *multi-aspect*, memberikan visualisasi hasil yang jelas, dan menjembatani gap antara model machine learning kompleks dengan kebutuhan stakeholder dalam menganalisis *feedback* pengguna secara terstruktur dan terukur.
5. Analisis menggunakan model yang dikembangkan berhasil mengidentifikasi aspek layanan yang dominan menerima sentimen positif dan negatif. Pada ketiga aplikasi (Mobile JKN, M-Paspor, dan SIGNAL), aspek *Reliability* (keandalan) secara konsisten menjadi sumber keluhan utama dan paling banyak menerima sentimen negatif dari pengguna. Sebaliknya aspek *Usability* (kemudahan penggunaan) menjadi kekuatan utama yang paling banyak menerima sentimen positif.

## 5.2 Saran

Sebagai tindak lanjut dari kesimpulan di atas, terdapat beberapa rekomendasi yang dapat menjadi pertimbangan untuk studi lanjutan terkait ABSA:

1. Penelitian selanjutnya dapat memperluas kategori aspek yang dianalisis. Selain tiga aspek yang telah ditentukan (*Usability, Reliability, Efficiency*), dapat ditambahkan aspek lain seperti *Security* (keamanan), *Features* (fitur), atau *Customer Service* (layanan pelanggan) untuk mendapatkan wawasan yang lebih komprehensif dan spesifik.
2. Disarankan untuk menggunakan dataset yang lebih besar dan beragam,
3. Untuk validasi performa yang lebih kokoh, penelitian mendatang dapat melakukan perbandingan secara langsung dengan model *baseline* lain seperti SVM atau arsitektur *transformer* yang berbeda.

## DAFTAR PUSTAKA

- Adimanggala, D., Bachtiar, F. A., & Setiawan, E. (2021). Evaluasi Topik Tersembunyi Berdasarkan Aspect Extraction menggunakan Pengembangan Latent Dirichlet Allocation. *Jurnal RESTI*, 5(3), 511–519. <https://doi.org/10.29207/resti.v5i3.3075>
- Amien, M. (2023). *Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan tentang sejarah, perkembangan teknologi, dan aplikasi NLP dalam bahasa Indonesia*. <http://arxiv.org/abs/2304.02746>
- Amien, S., Perdana, P., Bharata Aji, T., & Ferdiana, R. (2021). Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia (Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset). Dalam *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 10, Nomor 3).
- Badan Pusat Statistik. (2024). *Statistik Telekomunikasi Indonesia 2023*.
- Chandradev, V., Made, I., Dwi Suarjaya, A., Putu, I., & Bayupati, A. (2023). *Chandradev, Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT*.
- Chinnasamy, P., Suresh, V., Ramprathap, K., Jebamani, B. J. A., Srinivas Rao, K., & Shiva Kranthi, M. (2022). COVID-19 vaccine sentiment analysis using public opinions on Twitter. *Materials Today: Proceedings*, 64, 448–451. <https://doi.org/10.1016/j.matpr.2022.04.809>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>
- Fehle, J., Donhauser, N., Kruschwitz, U., Hellwig, N. C., & Wolff, C. (2025). *German Aspect-based Sentiment Analysis in the Wild: B2B Dataset Creation and Cross-Domain Evaluation*. <https://github.com/JakobFehle/>

- Filemon Haganta Kaban, A., & Yudistira, N. (2021). *Analisis Sentimen Aplikasi E-Government berdasarkan Ulasan Pengguna menggunakan Metode Maximum Entropy dan Seleksi Fitur Mutual Information* (Vol. 5, Nomor 4). <http://j-ptiik.ub.ac.id>
- Fitriyah, N., Warsito, B., Asih, D., & Maruddani, I. (2020). ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM). *JURNAL GAUSSIAN*, 9(3), 376–390. <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Hakim, G., Fatyanosa, T. N., & Widodo, A. W. (2024). *Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT* (Vol. 8, Nomor 10). <http://j-ptiik.ub.ac.id>
- Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). *Transformer in Transformer*. <https://gitee.com/mindspore/models/>
- Hidayat, M. N., & Pramudita, R. (2024). Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT. *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS: Journal of Information Management*, 8(2), 161. <https://doi.org/10.51211/imbi.v8i2.2719>
- International Organization for Standardization. (2011). *ISO/IEC 25010:2011 — Systems and software quality requirements and evaluation (SQuaRE) — System and software quality models*.
- Iskandar, J. W., & Nataliani, Y. (2021). Comparison of Naïve Bayes, SVM, and k-NN for Aspect-Based Gadget Sentiment Analysis. *Jurnal RESTI*, 5(6), 1120–1126. <https://doi.org/10.29207/resti.v5i6.3588>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. <http://arxiv.org/abs/2011.00677>
- Merdiansah, R., & Ali Ridha, A. (2024). Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), 221–228.

- Mohiuddin, K., Alam, M. A., Alam, M. M., Welke, P., Martin, M., Lehmann, J., & Vahdati, S. (2023). Retention is All You Need. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 4752–4758. <https://doi.org/10.1145/3583780.3615497>
- Muliawaty, L., & Hendryawan, S. (2020). PERANAN E-GOVERNMENT DALAM PELAYANAN PUBLIK (STUDI KASUS: MAL PELAYANAN PUBLIK KABUPATEN SUMEDANG). *Jurnal Ilmu Administrasi*, 11(2).
- Naquitasia, R., Hatta Fudholi, D., & Iswari, L. (2022). *ANALISIS SENTIMEN BERBASIS ASPEK PADA WISATA HALAL DENGAN METODE DEEP LEARNING* (Vol. 16, Nomor 2). <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- Nurian, A., Ma'arif, M. S., Amalia, I. N., & Rozikin, C. (2024). ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAIVE BAYES CLASSIFIER. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(1). <https://doi.org/10.23960/jitet.v12i1.3631>
- Putri, R. R., & Cahyono, N. (2024). ANALISIS SENTIMEN KOMENTAR MASYARAKAT TERHADAP PELAYANAN PUBLIK PEMERINTAH DKI JAKARTA DENGAN ALGORITMA SUPER VECTOR MACHINE DAN NAIVE BAYES. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Nomor 2).
- Radiena, G., & Nugroho, A. (2023). ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI KAI ACCESS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE. Dalam *Jurnal Pendidikan Teknologi Informasi (JUKANTI)* (Nomor 1).
- Rahman, I. F., Hasanah, A. N., & Heryana, N. (2024). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SAMSAT DIGIITAL NASIONAL (SIGNAL) DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2). <https://doi.org/10.23960/jitet.v12i2.4073>

- Trușcă, M. M., & Frasincar, F. (2023). Survey on aspect detection for aspect-based sentiment analysis. *Artificial Intelligence Review*, 56(5), 3797–3846. <https://doi.org/10.1007/s10462-022-10252-y>
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (t.t.). *Attention Is All You Need*.
- Widiansyah, M., Frazna Az-Zahra, F., & Pambudi, A. (2024). Fine-Tuning Model Indobert (Indonesian Bidirectional *Encoder* Representations from Transformers) untuk Analisis Sentimen Berbasis Aspek pada Aplikasi M-Paspor. Dalam *Journal of Informatic Engineering (JOUTICA)*. <https://doi.org/10.30736/informatika.v9i2.1310>
- Zhou, J., Huang, J. X., Chen, Q., Hu, Q. V., Wang, T., & He, L. (2019). Deep learning for aspect-level sentiment classification: Survey, vision, and challenges. Dalam *IEEE Access* (Vol. 7, hlm. 78454–78483). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2019.2920075>

