

**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI  
TELEMEDICINE MENGGUNAKAN METODE  
FINE-TUNING INDOBERT**

**SKRIPSI**

**Oleh:**

**SITA MAULIDIA ALYATU ZAHRA**  
**NIM. 210605110068**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI  
TELEMEDICINE MENGGUNAKAN METODE  
*FINE-TUNING* INDOBERT**

**SKRIPSI**

Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :  
**SITA MAULIDIA ALYATU ZAHRA**  
**NIM. 210605110068**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

**HALAMAN PERSETUJUAN**

**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI  
TELEMEDICINE MENGGUNAKAN METODE  
FINE-TUNING INDOBERT**

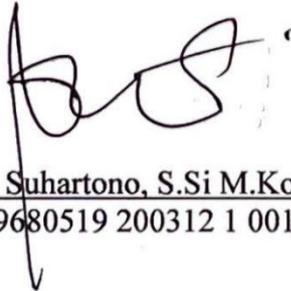
**SKRIPSI**

**Oleh :**

**SITA MAULIDIA ALYATU ZAHRA**  
**NIM. 210605110068**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 15 Juni 2025

Pembimbing I,



Prof. Dr. Suhartono, S.Si M.Kom  
NIP. 19680519 200312 1 001

Pembimbing II,



Dr. M. Imamudin Lc, MA  
NIP. 19740602 200901 1 010

Mengetahui,

Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. H. Fachrul Kurniawan, M.MT., IPU  
NIP. 19771020 200912 1 001

**HALAMAN PENGESAHAN**

**ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI  
TELEMEDICINE MENGGUNAKAN METODE  
FINE-TUNING INDOBERT**

**SKRIPSI**

**Oleh:**

**SITA MAULIDIA ALYATU ZAHRA**  
**NIM. 210605110068**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer ( S.Kom )  
Tanggal: 16 Juni 2025

**Susunan Dewan Penguji**

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom.  
NIP. 19761013 200604 1 004

Anggota Penguji I : Tri Mukti Lestari, M.Kom.  
NIP. 19911108 202012 2 005

Anggota Penguji II : Prof. Dr. Suhartono, M.Kom.  
NIP. 19680519 200312 1 001

Anggota Penguji III : Dr. M. Imamudin Lc, MA.  
NIP. 19740602 200901 1 010

(  
(  
(  
(  
(

Mengetahui dan Mengesahkan,  
Ketua Program Studi Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Ir. Fachrul Kurniawan, M.MT., IPU  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Sita Maulidia Alyatu Zahra  
NIM : 210605110068  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : Analisis Sentimen Pengguna Terhadap Aplikasi  
Telemedicine Menggunakan Metode Fine-Tuning  
Indobert.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 Juni 2025

Yang membuat pernyataan,



Sita Maulidia Alyatu Zahra

NIM.210605110068

## **MOTTO**

*Akan selalu ada jalan untuk setiap niat baik yang kita usahakan*

## HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan puji syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya, serta shalawat dan salam kepada Rasulullah SAW.

Berkat izin-Nya, skripsi ini dapat terselesaikan.

Persembahan tulus ini saya haturkan teristimewa kepada kedua orang tua tercinta, Abah Syamsun Hadi dan Ibu Rois Amaliyah. Kalian adalah pilar utama dalam hidup penulis, sumber cinta yang tak terhingga, doa yang tak pernah putus, pengorbanan yang tak ternilai, serta motivasi yang tak pernah padam di setiap langkah penulis.

Untuk seluruh keluarga besar yang senantiasa memberikan semangat dan kasih sayang. Kepada teman, sahabat, dan pasangan yang selalu ada, berbagi tawa dan duka, serta tak henti memberikan dukungan dan pengingat.

Dan terakhir, persembahan ini juga untuk diri saya sendiri, yang telah berjuang, bertahan, dan tidak menyerah dalam setiap prosesnya. Terima kasih atas segala ketekunan, kesabaran, dan semangat yang terus menyala. Semoga karya ini bermanfaat bagi ilmu pengetahuan dan sesama.

## KATA PENGANTAR

Kata Dengan segala kerendahan hati dan rasa syukur yang tiada terhingga, segala puji hanyalah milik Allah SWT, atas rahmat, hidayah, dan karunia-Nya yang tak terhingga, sehingga skripsi dengan judul "Analisis Sentimen Pengguna Terhadap Aplikasi Telemedicine Menggunakan Metode Fine-Tuning Indobert" ini dapat terselesaikan. Tanpa ridho dan pertolongan-Nya, mustahil karya ini dapat terwujud.

Selama proses penyelesaian skripsi ini, penulis menghadapi berbagai tantangan dan rintangan. Namun, berkat bimbingan, dukungan, bantuan, serta doa dari berbagai pihak, baik secara moral maupun material, segala hambatan tersebut dapat dilewati. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Prof. Dr. H.M. Zainuddin, MA., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Suhartono, M.Kom. selaku Dosen Pembimbing I dan Dr. M. Imamudin Lc, MA. selaku Dosen Pembimbing II, terima kasih atas waktu, arahan, bimbingan, serta kesabaran yang tak pernah lelah dalam

membimbing dan memberikan masukan berharga hingga skripsi ini dapat terselesaikan dengan baik.

5. Dr. M. Ainul Yaqin, M.Kom. selaku Dosen Penguji I dan Tri Mukti Lestari, M.Kom. selaku Dosen Penguji II, terima kasih atas kritik, saran, dukungan, dan wawasan yang telah diberikan, sangat membantu dalam menyempurnakan skripsi ini.
6. Segenap Dosen dan segenap staff Program Studi Teknik Informatika, terima kasih atas ilmu dan bimbingan yang telah diberikan selama masa perkuliahan ini.
7. Abah dan Ibuk tercinta, terima kasih atas doa yang tak pernah putus, dukungan moral dan material, serta motivasi yang menjadi sumber kekuatan terbesar dalam setiap langkah penulis.
8. Charis Maulana Setya Adi selaku teman, sahabat, dan pasangan yang telah membersamai, memberikan dukungan, motivasi, dan selalu ada untuk penulis.
9. Teman-teman di Surabaya, Icha, Retno, Kapal, Tiara, Dori, Balqis, Cinta, Ila, Dita yang telah mendukung dan mendoakan penulis walau terpaut jarak.
10. Teman-teman seperjuangan bimbingan anak Prof yang telah menemani dan menyemangati dalam langkah penyusunan skripsi ini.
11. Teman-teman Sri Rejeki dan Pak Tyo, yang telah memberikan memberikan doa, bantuan, dan dukungan yang sangat berarti bagi penulis.
12. Teman-teman Aster Teknik Teknik Informatika 2021 yang telah memberi dukungan dan bantuan dalam penyusunan skripsi ini.

13. Seluruh pihak yang telah membantu penulis dalam menyelesaikan skripsi ini, baik secara langsung maupun tidak langsung yang tidak dapat disebutkan satu per satu.
14. Dan terakhir, untuk diri sendiri, yang telah berjuang, bertahan, dan tidak menyerah dalam setiap prosesnya. Terima kasih atas ketekunan dan semangat yang terus menyala.

Malang, 15 Juni 2025

Penulis

## DAFTAR ISI

HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN .....	v
MOTTO .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL .....	xiv
ABSTRAK .....	xv
ABSTRACT .....	xvi
الملخص .....	xvii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	7
1.3 Batasan Masalah .....	7
1.4 Tujuan Penelitian .....	7
1.5 Manfaat Penelitian .....	7
<b>BAB II STUDI PUSTAKA .....</b>	<b>9</b>
2.1 Penelitian Terkait .....	9
2.2 Analisis Sentimen .....	14
2.3 Telemedicine.....	16
2.4 IndoBERT .....	17
2.5 <i>Fine-Tuning</i> IndoBERT .....	19
2.6 <i>Early Stopping</i> .....	22
<b>BAB III DESAIN DAN IMPLEMENTASI .....</b>	<b>24</b>
3.1 Desain Penelitian .....	24
3.2 Pengumpulan Data .....	25
3.3 Desain Sistem.....	28
3.3.1 <i>Cleaning Data</i> .....	29
3.3.2 <i>Preprocessing</i> .....	29
3.3.3 Pelatihan Model.....	34
3.3.4 <i>Fine-Tuning</i> IndoBERT.....	35
3.3.5 Evaluasi Model.....	41
3.4 Skenario Pengujian .....	42
<b>BAB IV UJI COBA DAN PEMBAHASAN.....</b>	<b>44</b>
4.1 Pengumpulan Data .....	44
4.2 Uji Coba.....	46
4.2.1 <i>Cleaning Data</i> .....	47
4.2.2 <i>Preprocessing</i> .....	48
4.2.3 Pelatihan Model.....	51
4.2.4 Pengujian Model <i>Fine-Tuning</i> IndoBERT .....	52
4.3 Hasil Uji Coba.....	54

4.3.1 Pengujian Skenario 1 .....	55
4.3.2 Pengujian Skenario 2 .....	56
4.4 Pembahasan.....	63
4.5 Visualisasi Kategori pada Ulasan .....	68
4.6 Visualisasi Kata pada Ulasan .....	70
4.7 Integrasi Islam.....	74
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>78</b>
5.1 Kesimpulan .....	78
5.2 Saran .....	79
<b>DAFTAR PUSTAKA</b>	

## DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian.....	24
Gambar 3.2 Desain Sistem.....	28
Gambar 4.1 Distribusi Sentimen Ulasan Pengguna.....	45
Gambar 4.2 Distribusi Data Ulasan.....	45
Gambar 4.3 Hasil Cleaning Data.....	47
Gambar 4.4 Visualisasi Perbandingan Learning rate.....	56
Gambar 4.5 Confusion Matrix pada Split Data 90:10.....	58
Gambar 4.6 Confusion Matrix pada Split Data 70:30.....	60
Gambar 4.7 Confusion Matrix pada Split Data 70:30.....	62
Gambar 4.8 Visualisasi berdasarkan validation loss dan validation accuracy.....	64
Gambar 4.9 Visualisasi perbandingan Confusion Matrix setiap Split Data.....	65
Gambar 4.10 Visualisasi hasil kategori pada.....	69
Gambar 4.11 WordCloud Sentimen Positif.....	70
Gambar 4.12 Top 5 Kata pada Sentimen Positif.....	71
Gambar 4.13 WordCloud Sentimen Negatif.....	72
Gambar 4.14 Top 5 Kata pada Sentimen Negatif.....	72
Gambar 4.15 WordCloud Sentimen Netral.....	73
Gambar 4.16 Top 5 Kata pada Sentimen Netral.....	73

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait .....	12
Tabel 3.1 Detail sample data .....	26
Tabel 3.2 Contoh Data Ulasan .....	27
Tabel 3.3 Proses Cleaning dan Case Folding .....	30
Tabel 3.4 Proses Normalisasi .....	32
Tabel 3.5 Proses Tokenizing .....	33
Tabel 3.6 Parameter Fine-Tuning IndoBERT .....	36
Tabel 3.7 Nilai Logits .....	38
Tabel 3.8 Confusion Matrix dalam analisis sentimen .....	41
Tabel 3.9 Skenario Pengujian .....	43
Tabel 4.1 Contoh Dataset .....	46
Tabel 4.2 Hasil Case Folding dan Text Cleaning .....	49
Tabel 4.3 Hasil Normalisasi .....	50
Tabel 4.4 Hasil Tokenisasi .....	50
Tabel 4.5 Hasil Stopword Removal .....	51
Tabel 4.6 Data Split .....	52
Tabel 4.7 Perbandingan Learning rate .....	55
Tabel 4.8 Hasil Early Stopping pada split data 90:10 .....	57
Tabel 4.9 Hasil perhitungan Confusion Matrix pada Split Data 90:10 .....	58
Tabel 4.10 Hasil Early Stopping pada split data 80:20 .....	59
Tabel 4.11 Hasil perhitungan Confusion Matrix pada Split Data 80:20 .....	60
Tabel 4.12 Hasil Early Stopping pada split data 70:30 .....	61
Tabel 4.13 Hasil perhitungan Confusion Matrix pada Split Data 70:30 .....	62
Tabel 4.14 Hasil Rata-Rata Confusion Matrix pada setiap rasio .....	65
Tabel 4.15 Hasil pengkategorian ulasan .....	69

## ABSTRAK

Zahra, Sita Maulidia Alyatu. 2025. **Analisis Sentimen Pengguna Terhadap Aplikasi Telemedicine Menggunakan Metode Fine-Tuning Indobert**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Suhartono, M.Kom. (II) Dr. M. Imamudin Lc, MA.

**Kata kunci:** Analisis Sentimen, IndoBERT, Telemedicine, *Hyperparameter*, *Early Stopping*

Perkembangan pesat teknologi informasi dan komunikasi telah membawa telemedicine menjadi inovasi krusial dalam layanan kesehatan di Indonesia, terutama pasca-pandemi COVID-19, namun adopsi luasnya masih menghadapi tantangan terkait persepsi pengguna. Studi empiris ini secara sistematis menginvestigasi optimasi model *Fine-tuning* IndoBERT untuk klasifikasi sentimen pada 7.000 ulasan pengguna aplikasi telemedicine halodoc dan alodokter yang diperoleh dari *Elsevier Mendeley Data Repository*. Penelitian ini berfokus pada identifikasi *hyperparameter* optimal dan teknik pelatihan paling efektif guna meningkatkan akurasi dan mencegah *overfitting*. Implementasi *grid search* pada skenario 1 secara logis menentukan *learning rate*  $3e-05$  sebagai yang paling optimal, menghasilkan akurasi validasi 82.93% dan loss 0.4540, yang secara relevan mengungguli konfigurasi lain. Selanjutnya, Skenario 2 membuktikan efektivitas *early stopping* dengan *patience 5 epoch* dalam mencegah *overfitting*, memungkinkan model menghentikan pelatihan secara tepat. Perbandingan rasio pembagian data pada skenario 2 menunjukkan bahwa rasio 90:10 memberikan performa superior dengan akurasi validasi tertinggi 83.71%, *precision* 83.27%, *recall* 83.41%, dan *F1-score* 83.28%, menekankan pentingnya proporsi data yang seimbang berdasarkan teori *machine learning*. Secara keseluruhan, kombinasi optimasi *learning rate*, pembagian data 90:10, dan *early stopping* terbukti efektif dalam menghasilkan model *fine-tuning* IndoBERT yang tangguh dan akurat untuk analisis sentimen ulasan aplikasi telemedicine berbahasa Indonesia

## ABSTRACT

Zahra, Sita Maulidia Alyatu. 2025. **Analisis Sentimen Pengguna Terhadap Aplikasi Telemedicine Menggunakan Metode Fine-Tuning Indobert**. Thesis. Departemen of Informatics Engineering, Faculty of Science and Technology, State Islamic University Islamic Maulana Malik Ibrahim Malang. Supervisor: (I) Prof. Dr. Suhartono, M.Kom. (II) Dr. M. Imamudin Lc, MA.

**Keywords:** Sentiment Analysis, IndoBERT, Telemedicine, *Hyperparameter*, *Early Stopping*

The rapid development of information and communication technology has made telemedicine a crucial innovation in Indonesian healthcare, especially post-pandemic. However, its widespread adoption still faces challenges related to user perception. This empirical study systematically investigates optimizing the IndoBERT fine-tuning model for sentiment classification using 7,000 user reviews of Halodoc and Alodokter telemedicine apps obtained from the Elsevier Mendeley Data Repository. The study focuses on identifying optimal hyperparameters and the most effective training technique to improve accuracy and prevent overfitting. Scenario 1's grid search implementation logically determined a learning rate of  $3e-05$  to be optimal, achieving a validation accuracy of 82.93% and a loss of 0.4540. This configuration notably outperformed others. Furthermore, scenario 2 demonstrated the effectiveness of early stopping with patience set to five epochs in preventing overfitting and enabling the model to appropriately stop training. A comparison of the data-sharing ratio in Scenario 2 shows that the 90:10 ratio provides the best performance, with a validation accuracy of 83.71%, a precision of 83.27%, a recall of 83.41%, and an F1 score of 83.28%. This emphasizes the importance of balancing the proportion of data based on machine learning theory. In summary, the combination of learning rate optimization, a 90:10 data split, and early stopping was effective in creating a robust and accurate IndoBERT model for sentiment analysis of Indonesian telemedicine app reviews.

## الملخص

زهرة، ست مولديا عليّة. 2025. تحليل مشاعر المستخدمين تجاه تطبيقات التّطبيب عن بُعد باستخدام طريقة إندوبيرت للضبط الدقيق. الرسالة الجامعة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم مالانج الحكومية الإسلامية. المشرف: (I) الأستاذ الدكتور سوهارتونو، م. كوم (II). الدكتور. محمد. إمام الدين، ماجستير.

الكلمات الرئيسية: تحليل المشاعر، إندوبيرت، التّطبيب عن بُعد، المقياس الفائق، التوقف المبكر

لقد جعل التطور السريع لتكنولوجيا المعلومات والاتصالات من التّطبيب عن بُعد ابتكارًا حاسمًا في مجال الرعاية الصحية الإندونيسية، خاصة بعد الجائحة. ومع ذلك، لا يزال اعتمادها على نطاق واسع يواجه تحديات تتعلق بإدراك المستخدم. تبحث هذه الدراسة التجريبية بشكل منهجي في تحسين نموذج إندوبيرت للضبط الدقيق لتصنيف المشاعر باستخدام 7000 تقييم مستخدم لتطبيقات هالودوك وألودوكتر للتّطبيب عن بُعد التي تم الحصول عليها من مستودع بيانات إلسيفير منديلي. وتركز الدراسة على تحديد المعلومات الفائقة المثلى وتقنية التدريب الأكثر فعالية لتحسين الدقة ومنع الإفراط في التركيب. وقد حدد تطبيق البحث الشبكي للسيناريو 1 منطقيًا معدل تعلم قدره 0.05-3 ليكون الأمثل، محققًا دقة تحقق بلغت 82.93% وخسارة قدرها 0.4540. تفوق هذا التكوين بشكل ملحوظ على الآخرين. علاوةً على ذلك، أظهر السيناريو 2 فعالية التوقف المبكر مع ضبط الصبر على خمس حقبات في منع الإفراط في التهيئة وتمكين النموذج من التوقف عن التدريب بشكل مناسب. تُظهر المقارنة بين نسبة مشاركة البيانات في السيناريو 2 أن نسبة 90:10 توفر أفضل أداء، مع دقة تحقق تبلغ 83.71%، ودقة تبلغ 83.27%، واستدعاء 83.41%، ودرجة F1 تبلغ 83.28%. وهذا يؤكد أهمية موازنة نسبة البيانات بناءً على نظرية التعلم الآلي. باختصار، كان الجمع بين تحسين معدل التعلم، وتقسيم البيانات بنسبة 90:10، والتوقف المبكر فعالاً في إنشاء نموذج إندوبيرت قوي ودقيق لتحليل المشاعر لمراجعات تطبيقات التّطبيب عن بُعد الإندونيسية.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kemajuan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai bidang, termasuk kesehatan, dengan telemedicine sebagai salah satu inovasi yang berkembang pesat. Telemedicine memungkinkan pasien untuk berkonsultasi dengan tenaga medis tanpa perlu tatap muka langsung, dan penggunaan layanan ini semakin meningkat di Indonesia, terutama setelah pandemi COVID-19. Data dari Kementerian Kesehatan menunjukkan adanya lonjakan pemanfaatan telemedicine selama pandemi, dengan sekitar 17,9 juta konsultasi medis dilakukan melalui 19 perusahaan telemedik pada 2022. Layanan ini menawarkan kemudahan konsultasi melalui berbagai platform, seperti video, telepon, atau chat, yang memungkinkan pasien memperoleh diagnosis awal, rekomendasi pengobatan, dan resep obat tanpa perlu datang langsung ke fasilitas kesehatan. Selain menghemat waktu dan biaya, telemedicine juga membantu mengurangi beban rumah sakit yang sering mengalami lonjakan pasien.

Meskipun pemanfaatan telemedicine terus berkembang, terdapat beberapa tantangan dalam adopsi layanan ini di Indonesia, terutama terkait dengan persepsi dan sentimen masyarakat. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2022, sekitar 95,11% masyarakat Indonesia belum memanfaatkan layanan telemedicine, dengan alasan utama adalah kenyamanan berkonsultasi secara langsung. Survei dari Katadata Insight Center pada tahun yang sama menemukan

bahwa 72% responden lebih memilih konsultasi tatap muka karena merasa lebih percaya pada diagnosa yang diberikan secara langsung. Di sisi lain, 30,5% responden meragukan ketepatan diagnosis melalui telemedicine, sedangkan 14,9% mengidentifikasi kualitas jaringan internet sebagai masalah yang mengurangi minat mereka. Dengan lebih dari 95% populasi yang belum beralih ke telemedicine, terdapat resistensi signifikan terhadap penggunaan teknologi ini, yang mencerminkan ketidaknyamanan dan keraguan masyarakat dalam beralih dari sistem kesehatan tradisional ke layanan kesehatan digital.

Dalam perspektif Islam, menjaga kesehatan dan mencari pengobatan merupakan bagian dari ikhtiar yang dianjurkan. Sebagaimana disebutkan dalam Surah As-Syu'ara ayat 80, Allah SWT berfirman:

وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ ﴿٨٠﴾

*“Apabila aku sakit, Dialah yang menyembuhkanku” (Qs. Asy-Syu'ara: 80)*

Menurut tafsir ringkas Kementrian Agama RI pada website Quran

Kemenag menyatakan bahwa:

*“Dan apabila aku sakit, dialah pada hakikatnya yang menyembuhkan aku, baik melalui sebab atau tidak.”*

Dalam perspektif Islam, menjaga kesehatan dan berusaha untuk sembuh dari penyakit merupakan bagian dari upaya ikhtiar yang sangat dianjurkan. Allah SWT dalam Surah As-Syu'ara ayat 80 menegaskan bahwa Dia adalah sumber kesembuhan yang sejati, baik melalui sebab-sebab yang nampak maupun yang tidak tampak oleh manusia. Dalam ayat ini, mengandung makna bahwa meskipun ada

ikhtiar manusia melalui pengobatan dan perawatan, kesembuhan sejatinya berasal dari Allah.

Begitu juga sebagaimana yang disebutkan dalam Surah Ar-Ra'd ayat 11, Allah SWT berfirman:

لَهُ مَعْبُوتٌ مِّنْ بَيْنِ يَدَيْهِ وَمَنْ خَلْفَهُ ۖ يَحْفَظُونَهُ ۗ مِنْ أَمْرِ اللَّهِ إِنَّ اللَّهَ لَا يُعَيِّرُ مَا بِقَوْمٍ حَتَّىٰ يُعَيِّرُوهُمَا ۗ  
بِأَنفُسِهِمْ ۗ وَإِذَا أَرَادَ اللَّهُ بِقَوْمٍ سُوءًا فَلَا مَرَدَّ لَهُ ۗ وَمَا لَهُمْ مِنْ دُونِهِ مِنِّ وَّالٍ ﴿١١﴾

*“Baginya (manusia) ada (malaikat-malaikat) yang menyertainya secara bergiliran dari depan dan belakangnya yang menjaganya atas perintah Allah. Sesungguhnya Allah tidak mengubah keadaan suatu kaum hingga mereka mengubah apa yang ada pada diri mereka. Apabila Allah menghendaki keburukan terhadap suatu kaum, tidak ada yang dapat menolaknya, dan sekali-kali tidak ada pelindung bagi mereka selain Dia.” (QS. Ar-Ra’d: 11).*

Menurut Imam Baidhawi dalam kitab *tafsir Anwarut Tanzil wa Asrarut Ta’wil* pada website NU Online yang ditulis oleh Ahmad Mundzir (2019) menyatakan bahwa:

*“Sesungguhnya Allah tidak mengganti sesuatu yang ada pada kamu dari kesehatan dan kenikmatan sampai mereka mengubah dengan individu mereka dari keadaan yang baik dengan keadaan yang buruk.” (Al-Baidhawi, Anwarut Tanzil wa Asrarut Ta’wil, [Daru Ihyait Turats al-Arabi: Beirut), juz 3, hal. 183)*

Ayat ini mengandung prinsip dasar bahwa perubahan dalam kondisi seseorang tidak terjadi begitu saja, melainkan membutuhkan upaya dan kesadaran individu untuk mengambil langkah yang diperlukan. Dalam konteks kesehatan, hal ini menunjukkan bahwa kesembuhan tidak hanya bergantung pada keadaan, tetapi juga pada usaha pasien terutama dalam mencari pengobatan yang tepat. Oleh karena itu, keseimbangan antara usaha medis dan tawakal menjadi prinsip utama dalam menghadapi penyakit. Dengan kata lain, berobat bukan hanya sekadar tindakan

fisik, tetapi juga bagian dari keyakinan dan keimanan seseorang kepada Allah sebagai Sang Penyembuh.

Dalam era digital saat ini, telemedicine menjadi salah satu bentuk ikhtiar yang memungkinkan pasien untuk mengakses layanan kesehatan secara lebih cepat dan efisien, terutama bagi mereka yang memiliki keterbatasan geografis atau fisik. Pemanfaatan telemedicine mencerminkan prinsip bahwa manusia perlu mengambil langkah aktif dalam menjaga kesehatannya, sejalan dengan ajaran Islam yang mengajarkan bahwa usaha adalah bagian tak terpisahkan dari takdir.

Dalam konteks ini, analisis sentimen menjadi alat yang sangat bermanfaat untuk memahami bagaimana masyarakat menilai layanan telemedicine. Beberapa penelitian sebelumnya, seperti yang dilakukan oleh Sholikun (2021), telah menyoroti pentingnya analisis sentimen untuk mengukur kepuasan pengguna terhadap aplikasi telemedicine seperti Halodoc.

Dari sudut pandang akademis, analisis sentimen merupakan metode dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) dan pembelajaran mesin (*Machine Learning*) yang bertujuan untuk mengklasifikasikan opini masyarakat menjadi sentimen positif, negatif, atau netral. Teknik ini telah banyak diterapkan di berbagai bidang, seperti pemasaran, politik, dan layanan pelanggan, untuk memahami respons publik terhadap produk, kebijakan, atau layanan tertentu. Dalam konteks telemedicine, pendekatan ini dapat memberikan wawasan berharga mengenai bagaimana masyarakat menilai kualitas layanan kesehatan digital.

Meskipun analisis sentimen telah banyak diterapkan di berbagai bidang, penerapannya dalam konteks telemedicine di Indonesia masih terbatas,

menciptakan celah yang perlu diisi melalui penelitian lebih lanjut. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan menganalisis sentimen terhadap aplikasi telemedicine populer di Indonesia. Sebelumnya, Rufaida (2023) telah melakukan analisis opini masyarakat terhadap platform telemedicine dengan menggabungkan metode kamus sentimen (InSet dan Sentiwords) dan *Support Vector Machine* (SVM). Hasilnya menunjukkan bahwa metode InSet berhasil mencapai akurasi 86,24% dalam analisis ulasan Halodoc dan 84,28% untuk Alodokter.

Berbeda dengan penelitian sebelumnya, penelitian ini akan menggunakan pendekatan yang lebih canggih, yaitu model NLP berbasis IndoBERT. IndoBERT adalah model pra-pelatihan berbasis Transformers yang dikembangkan khusus untuk bahasa Indonesia. Dengan melakukan *fine-tuning* terhadap dataset spesifik, model ini dapat lebih disesuaikan dengan karakteristik data yang dianalisis, sehingga menghasilkan prediksi yang lebih akurat. Penelitian oleh Jayadianti dkk. (2022) menunjukkan bahwa penggunaan model berbasis Transformers dalam analisis sentimen dapat meningkatkan akurasi dibandingkan dengan metode tradisional.

Selain peningkatan dalam metode analisis, penelitian ini juga akan mengimplementasikan *optimizer* AdamW dan *early stopping*. AdamW adalah algoritma optimisasi yang menggabungkan keuntungan dari algoritma Adam dan penyesuaian bobot regularisasi (*weight decay*), yang membantu model untuk lebih efisien dalam pembelajaran dan meningkatkan kinerja prediksi. Penelitian oleh Mahajaya dkk. (2024) menunjukkan bahwa penggunaan *optimizer* AdamW dapat

meningkatkan efisiensi pelatihan model dalam klasifikasi penyakit paru-paru. Di sisi lain, *early stopping* digunakan untuk mencegah *overfitting* dengan menghentikan pelatihan model ketika kinerja pada data validasi tidak lagi membaik, sehingga membantu model untuk generalisasi dengan lebih baik. Penelitian yang dilakukan oleh Anam dkk (2023) menunjukkan bahwa penerapan *early stopping* dapat meningkatkan kinerja model dalam klasifikasi teks dengan mencegah *overfitting* dan mempercepat konvergensi pelatihan.

Selain peningkatan dalam metode analisis, penelitian ini juga akan mencakup dua aplikasi telemedicine utama yang paling populer di Indonesia, yaitu aplikasi Halodoc dan Alodokter. Berdasarkan komparasi yang dirilis oleh *Top Brand 2024*, Halodoc menjadi aplikasi telemedicine dengan jumlah pengguna tertinggi, dipilih oleh 60,50% responden, diikuti oleh Alodokter (19,20%). Dengan menganalisis kedua aplikasi ini, penelitian dapat menggambarkan variasi sentimen masyarakat antar platform.

Hasil penelitian ini diharapkan memberi kontribusi baik secara akademik maupun praktis. Bagi pengembang aplikasi, hasilnya dapat memberikan wawasan untuk meningkatkan pengalaman pengguna, serta membantu pengembangan layanan telemedicine yang lebih efektif dalam meningkatkan akses dan kualitas kesehatan di Indonesia. Dengan demikian, penelitian ini penting untuk memperkaya pengetahuan di bidang teknologi informasi dan kesehatan, serta berkontribusi pada peningkatan kualitas hidup masyarakat Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka pernyataan masalah pada penelitian ini adalah bagaimana penerapan metode *fine-tuning* IndoBERT dengan teknik *early stopping* dapat mempengaruhi kinerja model?

## 1.3 Batasan Masalah

1. Penelitian terbatas pada ulasan pengguna dalam lingkup aplikasi telemedicine yang tersedia di platform *Google Play Store*, yaitu aplikasi Halodoc dan Alodokter.
2. Data ulasan yang digunakan terbatas sebanyak 7.000 ulasan dari kedua aplikasi, yaitu Halodoc dan Alodokter.
3. Data yang digunakan dalam penelitian ini hanya berupa ulasan teks yang berisi komentar pengguna tanpa rating
4. Penelitian ini hanya menggunakan ulasan yang ditulis dalam bahasa Indonesia.

## 1.4 Tujuan Penelitian

Menganalisis sentimen pengguna terhadap aplikasi telemedicine di Indonesia melalui penerapan metode *fine-tuning* pada model IndoBERT.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dengan menerapkan *fine-tuning* pada model IndoBERT untuk menganalisis sentiment pengguna aplikasi telemedicine berbahasa Indonesia, sehingga meningkatkan pemahaman terhadap

persepsi dan kebutuhan pengguna guna meningkatkan kualitas layanan telemedicine sesuai dengan kebutuhan dan harapan pengguna, sehingga akses terhadap layanan kesehatan menjadi lebih baik dan efisien.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 Penelitian Terkait**

Dilihat beberapa tahun terakhir, perkembangan teknologi di sektor kesehatan semakin pesat, dengan telemedicine menjadi salah satu inovasi paling berpengaruh di Indonesia. Sejak pandemi COVID-19, kebutuhan akan layanan kesehatan berbasis digital meningkat secara signifikan, mendorong masyarakat untuk lebih terbuka terhadap konsultasi medis jarak jauh. Aplikasi seperti Halodoc dan Alodokter menjadi pilihan utama dalam memberikan layanan kesehatan yang praktis, efisien, dan dapat diakses kapan saja tanpa harus melakukan pertemuan langsung dengan tenaga medis. Namun, meskipun teknologi ini menawarkan berbagai kemudahan, pemahaman mendalam mengenai persepsi dan kepuasan pengguna menjadi aspek krusial untuk memastikan layanan yang diberikan sesuai dengan harapan masyarakat. Dalam hal ini, analisis sentimen terhadap ulasan pengguna menjadi pendekatan yang sangat efektif untuk mengidentifikasi aspek positif dan negatif dari layanan telemedicine serta memberikan wawasan bagi pengembang dalam meningkatkan kualitas layanan mereka.

Sejumlah penelitian antara tahun 2020 hingga 2024 telah dilakukan untuk menganalisis sentimen serta pemodelan topik dalam aplikasi telemedicine di Indonesia. Kajian-kajian ini menggunakan berbagai metode berbasis kecerdasan buatan dan pemrosesan bahasa alami (*Natural Language Processing/NLP*) untuk

mengeksplorasi opini pengguna secara lebih sistematis. Beberapa penelitian yang relevan dalam konteks ini antara lain:

Penelitian yang dilakukan oleh Siti Mutmainah dengan judul Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine pada *Google Play* Menggunakan BiLSTM dan LDA (2023) berfokus pada analisis sentimen dan pemodelan topik terhadap ulasan pengguna aplikasi telemedicine seperti Yesdok, SehatQ, KlikDokter, Halodoc, dan Alodokter, yang dikumpulkan dari Google Play Store. Hasil dari emuan ini memberikan wawasan penting mengenai persepsi masyarakat terhadap layanan telemedicine, serta memungkinkan pengembang untuk lebih memahami kebutuhan pengguna dalam meningkatkan kualitas aplikasi mereka.

Andhika Nurriszky Alfarobby dan Herry Irawan (2024) meneliti analisis sentimen dan pemodelan topik dalam ulasan pengguna aplikasi Gojek dan Grab di Google Play Store. Dengan judul Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online pada Ulasan Google Play Store Menggunakan IndoBERT dan BERTopic. Peneliti menggunakan IndoBERT untuk analisis sentimen dan BERTopic untuk pemodelan topik. Analisis topik dengan BERTopic mengungkapkan bahwa pengguna Gojek lebih sering mengeluhkan pengalaman dan kualitas layanan, sementara pengguna Grab lebih menyoroti akurasi titik penjemputan. Studi ini menunjukkan bagaimana pendekatan BERTopic dapat membantu mengelompokkan tema utama dalam opini pengguna secara lebih akurat, sehingga metode ini juga berpotensi diterapkan dalam analisis sentimen layanan telemedicine.

Penelitian dengan judul Analisis Sentimen Aplikasi Telemedicine di Twitter Menggunakan Metode Berbasis *Leksikon* dan *Naive Bayes Classifier* yang dilakukan oleh Arid Hasan, dkk. (2023), analisis sentimen diterapkan pada opini masyarakat terhadap aplikasi telemedicine yang diunggah di Twitter. Hasilnya menunjukkan bahwa metode Lexicon-Based dan Naïve Bayes ini mampu mengidentifikasi sentimen positif, negatif dan netral dari masyarakat, sehingga dapat menjadi dasar bagi penyedia layanan telemedicine untuk melakukan perbaikan dalam sistem pelayanan mereka. Selain itu, penelitian ini menunjukkan bahwa media sosial seperti Twitter merupakan sumber data yang berharga dalam memahami persepsi publik secara real-time terhadap layanan kesehatan digital.

Penelitian yang dilakukan oleh Lenggo Geni, dkk. (2023) yang melakukan Analisis Sentimen Tweet Menjelang Pemilu 2024 di Indonesia Menggunakan Model IndoBERT untuk menganalisis opini masyarakat Indonesia terhadap Pemilu 2024. Meskipun fokus penelitian ini berada dalam domain politik, metode yang digunakan relevan dalam konteks lain. Hasil penelitian menunjukkan bahwa model IndoBERT mampu mengklasifikasikan sentimen dengan tingkat akurasi 83,5%, membuktikan keunggulannya dalam memahami konteks bahasa Indonesia. Pendekatan ini dapat diadaptasi untuk menganalisis sentimen pengguna aplikasi telemedicine, mengingat kompleksitas bahasa yang digunakan dalam ulasan digital memerlukan model pemrosesan bahasa alami yang canggih.

Reynalda Nabila Cikania (2023) dalam penelitiannya yang berjudul Analisis Sentimen Review Pengguna Layanan Telemedicine Menggunakan *Naive Bayes* dan *Support Vector Machine* yang mengkaji sentimen pengguna terhadap layanan

Halodoc dan Alodokter berdasarkan ulasan yang diambil dari *Google Play Store*. Penelitian ini menggunakan dua metode utama, yaitu *Naive Bayes* dan *Support Vector Machine* (SVM) untuk klasifikasi sentimen. Hasilnya menunjukkan bahwa sentimen positif lebih dominan dibandingkan dengan sentimen negatif, dengan tingkat akurasi 87,67% untuk model SVM. Penelitian ini memberikan wawasan mengenai aspek-aspek layanan yang mendapat apresiasi tinggi dari pengguna serta area yang masih memerlukan perbaikan.

Berdasarkan studi penelitian terkait yang disampaikan sebelumnya, disusun tabel 2.1 untuk membandingkan setiap penelitian tersebut.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Metode	Variabel Independen (Input)	Variabel Dependen (Output)	Hasil
1.	(Mutmainah dkk., 2023)	Metode Bilstm dan LDA	Ulasan pengguna aplikasi telemedicine	Sentimen (positif, netral, negatif), Topik (positif, netral, negatif)	Penelitian ini menggunakan 244.098 data ulasan. Dengan hasil sentimen positif sebanyak 45,9%, netral 14,3%, dan negatif 39,8%. Unruk Model BiLSTM mencapai akurasi 95%, sementara pemodelan topik menggunakan LDA (Latent Dirichlet Allocation) menghasilkan nilai koherensi terbaik pada topik positif (0,6437), netral (0,6296), dan negatif (0,6132).
2.	(Alfarobby, 2023)	IndoBERT dan Topic Modeling	Ulasan pengguna aplikasi Gojek dan Grab	Sentimen (positif, negatif), Topik ketidakpuasan	Dari 7.000 ulasan Gojek dan Grab, Gojek memiliki 60,8% ulasan positif dan 39,2% negatif, sementara Grab memiliki 79,5% positif dan 20,5% negatif. Dengan Akurasi model sebesar 95%.
3.	(Hasan dkk., 2023)	Lexicon-Based dan Naïve Bayes	Tweet terkait aplikasi telemedicine	Sentimen (positif, netral, negatif)	Menganalisis 1.248 tweet, ditemukan 908 positif, 168 netral, dan 172 negatif. Naïve Bayes mencapai akurasi 74%, dengan <i>precision</i> 98% dan

No	Nama Peneliti	Metode	Variabel Independen (Input)	Variabel Dependen (Output)	Hasil
					<i>recall</i> 86%, menunjukkan telemedicine mendapat respons positif dari masyarakat
4.	(Geni dkk., 2023)	IndoBERT	Tweet terkait Pemilu 2024	Sentimen (positif, netral, negatif)	Menggunakan 1.000 data tweet terkait Pemilu 2024, hasil sentimen menunjukkan 83,7% netral. Model IndoBERT-Large-P1 mencapai akurasi 83,5%, lebih baik dibandingkan TextBlob, Naïve Bayes, dan SVM dalam klasifikasi sentimen politik
5.	(Cikania, 2021)	Naïve Bayes dan SVM	Ulasan pengguna Halodoc	Sentimen (positif, negatif)	Dari 5.687 ulasan Halodoc, 87,67% positif dan 12,33% negatif. Naïve Bayes memiliki akurasi 87,77%, sedangkan SVM Kernel RBF lebih unggul dengan akurasi 86,1%, menunjukkan SVM lebih baik dalam klasifikasi sentimen pengguna
6.	(Imaduddin dkk., 2023)	<i>Fine-tuning</i> IndoBERT	Ulasan pengguna aplikasi kesehatan (Halodoc, Alodokter, KlikDokter)	Sentimen (positif, netral, negatif)	Penelitian ini mengembangkan sistem analisis sentimen untuk menilai kepuasan pengguna terhadap aplikasi layanan kesehatan di Indonesia, seperti Halodoc, Alodokter, dan KlikDokter. Dengan menggunakan dataset 9.310 ulasan pengguna, model ini mencapai akurasi 96%, presisi 95%, <i>recall</i> 96%, dan <i>F1-score</i> 95%, menunjukkan efektivitas IndoBERT dalam memahami konteks lokal bahasa Indonesia.
7.	(Putri dkk., 2025).	BERT	Ulasan pengguna aplikasi Segari	Sentimen (positif, negatif, netral)	Penelitian ini menerapkan metode BERT untuk analisis sentimen pada ulasan pengguna aplikasi Segari di Google Play Store. Menggunakan dataset ulasan pengguna,

No	Nama Peneliti	Metode	Variabel Independen (Input)	Variabel Dependen (Output)	Hasil
					model BERT berhasil mencapai akurasi 89% dalam mengklasifikasikan sentimen menjadi positif, negatif, dan netral. Presisi tertinggi dicapai pada sentimen positif dengan 91%, sedangkan presisi sentimen netral berada pada angka 69%, menunjukkan bahwa model ini cukup efektif dalam menganalisis sentimen pengguna pada aplikasi quick commerce.
8.	(Fatharani dkk., 2022)	Capsule Network	Percakapan chatbot berbahasa Indonesia	Intensi percakapan	Penelitian ini menggunakan model Capsule Network untuk mendeteksi intensi pada chatbot berbahasa Indonesia. Dengan menggunakan dataset percakapan chatbot, model BERT berhasil mencapai akurasi 94,5% dalam mengklasifikasikan intensi pengguna. Teknik <i>early stopping</i> diterapkan untuk mencegah <i>overfitting</i> , dan model ini terbukti efektif dalam mengenali berbagai intensi dalam percakapan, menjadikannya solusi yang andal untuk meningkatkan kinerja chatbot berbahasa Indonesia.
9.	Penelitian yang dilakukan	IndoBERT + Rules Set ( <i>Optimizer Adam W dan Early stopping</i> )	Ulasan pengguna aplikasi telemedicine	Sentimen (positif, netral, negatif)	

## 2.2 Analisis Sentimen

Analisis sentimen adalah proses untuk mengidentifikasi dan mengekstrak opini atau perasaan yang terkandung dalam teks, seperti ulasan produk, komentar

di media sosial, atau artikel berita. Tujuan utamanya adalah untuk memahami sikap, pendapat, dan emosi yang disampaikan oleh penulis mengenai topik tertentu. Dengan menganalisis sentimen, teks dapat dikategorikan ke dalam beberapa label, seperti positif, negatif, atau netral, yang memberikan gambaran yang lebih jelas mengenai bagaimana masyarakat memandang suatu isu atau produk (Mahira dkk., 2023).

Analisis sentimen memainkan peran penting dalam berbagai bidang, memberikan berbagai manfaat yang dapat digunakan untuk mengambil keputusan yang lebih tepat. Dalam dunia bisnis, analisis sentimen digunakan untuk mengawasi opini publik mengenai produk atau layanan yang ditawarkan. Hal ini memungkinkan perusahaan untuk mengevaluasi tingkat kepuasan pelanggan serta mengidentifikasi area yang perlu diperbaiki. Di sektor kesehatan, analisis ini dapat menggali persepsi masyarakat terhadap berbagai isu kesehatan yang berkembang. Oleh karena itu, analisis sentimen menjadi alat yang sangat penting dan berharga dalam pengambilan keputusan strategis di berbagai sektor (Laurensz & Sedyono, 2021).

Secara keseluruhan, analisis sentimen sangat penting dalam mengubah data teks yang awalnya tidak terstruktur menjadi informasi yang dapat ditindaklanjuti. Hal ini memungkinkan pihak yang berkepentingan untuk mengambil keputusan yang lebih informasional dan strategis. Dengan memahami sentimen publik terhadap produk atau layanan mereka, mereka dapat menyesuaikan strategi untuk lebih baik agar dapat memenuhi kebutuhan dan keinginan konsumen, sehingga

dapat memperbaiki hubungan dengan pelanggan dan meningkatkan daya saing di pasar.

### **2.3 Telemedicine**

Telemedicine adalah layanan kesehatan yang memanfaatkan teknologi informasi dan komunikasi, memungkinkan pasien dan tenaga medis berinteraksi tanpa kendala jarak. Tujuan utama dari telemedicine adalah untuk meningkatkan aksesibilitas dan efisiensi layanan kesehatan, terutama bagi mereka yang tinggal di daerah terpencil atau yang memiliki keterbatasan mobilitas. Dengan menggunakan telemedicine, pasien dapat menerima konsultasi medis, diagnosis, dan pengobatan dari jarak jauh, yang secara signifikan mengurangi kebutuhan perjalanan jauh ke fasilitas kesehatan. Selain itu, telemedicine juga memungkinkan pemantauan kesehatan secara terus-menerus, sehingga meningkatkan kualitas perawatan bagi pasien (Lestari & Gozali, 2021).

Di Indonesia, telemedicine telah menjadi solusi vital untuk mengatasi ketidakmerataan distribusi tenaga medis dan fasilitas kesehatan di berbagai wilayah. Selama pandemi COVID-19, misalnya, telemedicine memungkinkan pasien dengan kondisi kronis untuk terus mendapatkan perawatan tanpa harus mengunjungi rumah sakit guna mengurangi risiko penularan. Selain itu, telemedicine juga membantu untuk mengalihkan konsultasi rutin non-darurat ke *platform* digital. Meski demikian, implementasi telemedicine di Indonesia menghadapi sejumlah tantangan, seperti kurangnya infrastruktur teknologi yang merata, masalah terkait keamanan data, serta kebutuhan akan regulasi yang jelas untuk menjamin kualitas dan keamanan layanan. Oleh karena itu, pengembangan

telemedicine di Indonesia memerlukan kolaborasi antara pemerintah, penyedia layanan kesehatan, dan masyarakat untuk menciptakan ekosistem yang berkelanjutan dan mendukung (Bakhtiar, 2022).

## 2.4 IndoBERT

IndoBERT adalah model pemrosesan bahasa alami berbasis Transformer yang dikembangkan khusus untuk memahami teks dalam bahasa Indonesia. Sebagai versi adaptasi dari BERT (*Bidirectional Encoder Representations from Transformers*), IndoBERT dirancang untuk menangkap nuansa dan struktur unik bahasa Indonesia secara lebih akurat. Model ini hadir dalam beberapa varian, seperti IndoBERT-liteBase, IndoBERTBase, dan IndoBERTLarge, yang berbeda dalam kapasitas dan ukuran. Sebagai contoh, IndoBERTBase memiliki 12 lapisan tersembunyi, dengan 768 unit per lapisan, 12 kepala perhatian, serta lapisan feed-forward dengan dimensi 3.072. Model ini dilatih dengan menggunakan lebih dari 220 juta kata yang diperoleh dari berbagai sumber, seperti Wikipedia Indonesia, artikel dari Kompas, Tempo, Liputan6, dan korpus web Indonesia (Jayadianti dkk., 2022).

Dalam aplikasinya, IndoBERT digunakan untuk berbagai tugas pemrosesan bahasa alami (NLP) dalam bahasa Indonesia, seperti analisis sentimen, klasifikasi teks, dan deteksi entitas bernama. Penggunaan model ini melalui dua tahap utama: *pre-training* dan *fine-tuning*. Pada tahap *pre-training*, model dilatih untuk memahami struktur dan konteks bahasa Indonesia melalui dua tugas utama: *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Dalam MLM, sekitar 15% kata dalam setiap urutan token diganti dengan token [MASK], dan

model berusaha memprediksi kata yang hilang berdasarkan konteks sekitarnya. Sementara itu, NSP melatih model untuk memprediksi apakah suatu kalimat mengikuti kalimat sebelumnya. Setelah tahap pre-training, IndoBERT dapat disesuaikan melalui *fine-tuning* untuk tugas-tugas spesifik dengan menambahkan lapisan klasifikasi dan melatihnya menggunakan dataset yang relevan dengan tugas tersebut. Proses *fine-tuning* memungkinkan model untuk menyesuaikan parameter internalnya, sehingga menjadi lebih optimal untuk menyelesaikan tugas yang ditentukan (Putra, 2021).

IndoBERT memiliki berbagai keunggulan yang menjadikannya pilihan utama dalam analisis sentimen bahasa Indonesia. Model ini mampu memahami konteks bahasa Indonesia dengan mendalam, termasuk struktur kalimat dan nuansa khasnya. IndoBERT juga dapat menangani berbagai variasi bahasa, seperti slang, singkatan, dan istilah lokal yang sering digunakan dalam bahasa Indonesia. Kemudian model ini memiliki kemampuan transfer learning yang sangat baik, memungkinkan penerapannya untuk berbagai tugas NLP hanya dengan sedikit penyesuaian. IndoBERT juga dapat menangani teks panjang dengan sangat efisien berkat arsitektur Transformer yang digunakannya. Terakhir, model ini telah menunjukkan kinerja yang luar biasa dalam berbagai *benchmark* NLP bahasa Indonesia, menjadikannya lebih unggul dibandingkan model lain. Keunggulan-keunggulan tersebut menjadikan IndoBERT sebagai alat yang sangat efektif untuk analisis sentimen teks berbahasa Indonesia.

## 2.5 *Fine-Tuning* IndoBERT

Pada tahap *Fine-Tuning* IndoBERT, model dilatih untuk mempelajari dan menyesuaikan parameter internalnya berdasarkan data pelatihan yang telah diproses. Pada proses ini digunakan algoritma AdamW, yang merupakan varian dari *Adam optimizer* yang menggabungkan *weight decay* untuk regularisasi. Rumus untuk momentum pertama dan momentum kedua dalam AdamW (*Adaptive Moment Estimation with Weight Decay*) adalah algoritma optimisasi yang digunakan untuk memperbarui parameter model dalam pelatihan deep learning, seperti *fine-tuning* model IndoBERT. Algoritma ini merupakan varian dari Adam, yang mengkombinasikan dua jenis momentum (momentum pertama dan momentum kedua) untuk meningkatkan kecepatan konvergensi dan stabilitas selama pelatihan. AdamW mengatasi masalah *overfitting* dengan menambahkan *weight decay*, yang bertujuan untuk mengurangi ukuran parameter model dan mencegah model menjadi terlalu kompleks. Proses pembaruan parameter dalam AdamW dilakukan melalui tiga langkah utama: pembaruan momentum pertama (estimasi gradien), pembaruan momentum kedua (estimasi kuadrat gradien), dan pembaruan parameter itu sendiri.

Proses pertama yang dilakukan adalah menghitung momentum pertama ( $m_t$ ) dan momentum kedua ( $v_t$ ) untuk setiap parameter model berdasarkan gradien yang dihitung selama *backpropagation*. Seperti penelitian yang dilakukan oleh Mahajaya dkk., (2024) Momentum pertama adalah rata-rata eksponensial gradien, sedangkan momentum kedua adalah rata-rata eksponensial kuadrat gradien. Pembaruan momentum pertama dan momentum kedua dihitung sebagai berikut:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot gt \quad (2.1)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot gt^2 \quad (2.2)$$

Keterangan:

- $m_t$  : estimasi mometum pertama pada langkah  $t$
- $\beta_1$  : faktor *decay* momentum pertama
- $gt$  : gradien pada langkah  $t$
- $v_t$  : estimasi momentum kedua pada langkah  $t$
- $\beta_2$  : faktor *decay* momentum kedua

Dimana  $gt$  adalah gradien pada iterasi ke- $t$ ,  $\beta_1$  dan  $\beta_2$  adalah faktor *decay* untuk momentum pertama dan momentum kedua (biasanya 0.9 dan 0.999), kemudian  $m_t$  dan  $v_t$  adalah momentum yang diperbarui. Pembaruan ini memungkinkan model untuk lebih efisien, karena memanfaatkan informasi gradien dari iterasi sebelumnya.

Selanjutnya, koreksi bias diterapkan untuk memperbaiki estimasi momentum pada awal pelatihan, karena pada iterasi pertama  $m_t$  dan  $v_t$  memiliki nilai kecil yang terdistorsi. Koreksi bias dihitung dengan rumus berikut (Ramadhan & Hernadi, 2025):

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.3)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.4)$$

Keterangan:

- $\hat{m}_t$  : momentum pertama setelah koreksi bias
- $m_t$  : momentum pertama sebelum koreksi bias
- $\beta_1$  : faktor *decay* momentum pertama
- $t$  : iterasi (langkah ke- $t$ )
- $\hat{v}_t$  : Momentum kedua setelah koreksi bias
- $v_t$  : Momentum kedua sebelum koreksi bias
- $\beta_2$  : Faktor *decay* momentum kedua

Dengan bias correction, estimasi momentum menjadi lebih akurat, terutama pada iterasi awal. Setelah itu, pembaruan parameter model dilakukan dengan rumus AdamW, contohnya ketika parameter model pada iterasi ke-0 adalah  $\theta_0=1$ . Maka pembaruan parameter dihitung sebagai rumus berikut :

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta \cdot \lambda \cdot \theta_t \quad (2.5)$$

Keterangan:

- $\theta_{t+1}$  : parameter model yang diperbarui.
- $\theta_t$  : parameter model sebelumnya.
- $\eta$  : *learning rate*
- $\hat{m}_t$  : momentum pertama yang dikoreksi bias
- $\hat{v}_t$  : momentum kedua yang dikoreksi bias
- $\epsilon$  : konstanta kecil
- $\lambda$  : *weight decay*

Dalam implementasi model IndoBERT, AdamW memungkinkan model untuk memperbarui parameter secara efisien, mempercepat pelatihan, dan menghindari *overfitting*, menghasilkan model yang lebih stabil dan akurat dalam mengklasifikasikan sentimen ulasan pengguna aplikasi telemedicine.

Kemudian proses selanjutnya adalah *cross-entropy loss* yang berfungsi untuk menghitung perbedaan antara prediksi model dan label yang benar. Langkah selanjutnya dalam *fine-tuning* adalah proses *forward pass*, di mana input berupa teks yang telah melalui tokenisasi diubah menjadi representasi numerik (*logits*). *Logits* ini kemudian digunakan untuk menghitung probabilitas dengan menggunakan fungsi *Softmax* seperti pada penelitian yang dilakukan oleh (Masruroh dkk., 2023). Fungsi *Softmax* berfungsi mengubah skor mentah menjadi probabilitas kelas dengan rumus:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j^n e^{z_j}} \quad (2.6)$$

Keterangan:

$z_i$  : skor *logits* untuk kelas ke- $i$  (kelas positif, negatif, dan netral).

$\sum_j^n e^{z_j}$  : jumlah eksponensial dari semua *logits*

Setelah dilakukan perhitungan *softmax*, maka dilakukan perhitungan *cross-entropy loss* untuk mengukur ketidaksesuaian antara probabilitas prediksi model dan label sentimen yang sebenarnya. Berikut perhitungan dari *Cross-Entropy Loss* seperti penelitian yang dilakukan oleh (Pangestu dkk., 2020):

$$\mathcal{L} = - \sum_{c=1}^C y_c \log(p_c) \quad (2.7)$$

Keterangan:

$C$  : jumlah kelas (kelas: positif, negatif, dan netral).

$y_c$  : label aktual untuk kelas  $c$  (nilai biner, 1 jika benar, 0 jika tidak).

$p_c$  : probabilitas prediksi model untuk kelas  $c$  setelah *softmax*.

*Cross-entropy loss* ini memberikan ukuran seberapa jauh hasil prediksi model dari label yang benar, dan tujuannya adalah untuk meminimalkan nilai *loss* selama pelatihan.

## 2.6 *Early Stopping*

*Early stopping* adalah teknik regulasi dalam pelatihan model deep learning yang bertujuan mencegah *overfitting* dengan menghentikan proses pelatihan lebih awal sebelum model mulai menghafal data pelatihan. Teknik ini memantau kinerja model pada validation set dan menghentikan pelatihan ketika tidak ada peningkatan signifikan setelah sejumlah *epoch* tertentu, yang dikenal dengan parameter *patience*. Pemilihan nilai *patience* yang tepat sangat penting, nilai yang terlalu rendah dapat menyebabkan pelatihan berhenti sebelum model mencapai kinerja optimal, sementara nilai yang terlalu tinggi dapat menyebabkan *overfitting*. Secara

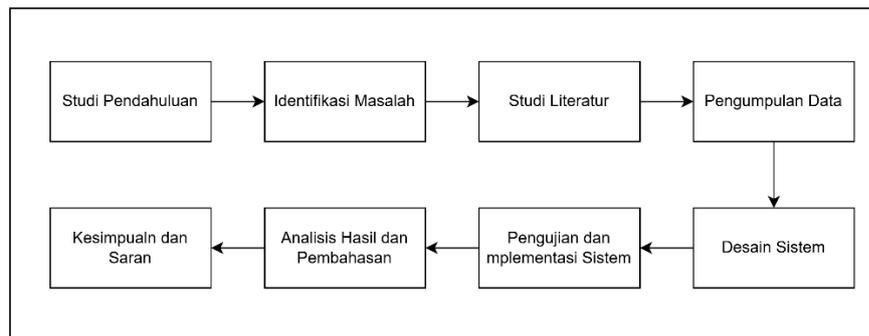
teori, *early stopping* bekerja dengan menghentikan pelatihan sebelum model mulai menghafal data pelatihan, sehingga mengurangi kompleksitas model dan meningkatkan kemampuannya untuk menggeneralisasi ke data yang belum pernah dilihat (Anam dkk., 2024).

Studi penelitian yang dilakukan oleh Anam, dkk (2024) dilakukan penerapan *early stopping* pada model CNN-LSTM untuk meningkatkan kinerja klasifikasi. Penelitian tersebut menemukan bahwa penerapan *early stopping* dapat meningkatkan akurasi model dalam klasifikasi data, dengan menghentikan pelatihan pada titik yang tepat sebelum model mulai mengalami *overfitting*. Temuan ini menunjukkan bahwa *early stopping* efektif dalam meningkatkan generalisasi model dan mencegah *overfitting*.

## BAB III

### DESAIN DAN IMPLEMENTASI

#### 3.1 Desain Penelitian



Gambar 3.1 Desain Penelitian

Berdasarkan Gambar 3.1 penelitian ini diawali dengan studi pendahuluan, di mana pada tahap ini dilakukan eksplorasi awal untuk memahami ruang lingkup permasalahan yang akan diteliti. Penelitian berlanjut ke tahap identifikasi masalah, yang berfungsi untuk merumuskan inti permasalahan yang akan diselesaikan. Selanjutnya, dilakukan studi literatur, yang bertujuan untuk menelaah penelitian-penelitian terdahulu yang relevan dengan topik yang diangkat. Kajian literatur ini memberikan landasan teoritis yang kuat, membantu dalam memahami metode yang telah digunakan sebelumnya, serta mengidentifikasi celah penelitian yang masih belum terjawab.

Setelah memperoleh pemahaman yang dari kajian literatur, penelitian memasuki tahap pengumpulan data, di mana data yang diperlukan diperoleh melalui *Elsevier Mendeley Data Repository* yang dapat diakses oleh publik. Data

yang terkumpul kemudian digunakan untuk membangun model atau sistem yang mendukung analisis yang dilakukan dalam penelitian ini.

Tahap berikutnya adalah perancangan sistem, yang mencakup pengembangan arsitektur serta mekanisme yang akan digunakan untuk mengolah data dan menjalankan analisis penelitian. Desain sistem ini disusun dengan mempertimbangkan efisiensi, akurasi, serta relevansi terhadap tujuan penelitian agar hasil yang diperoleh dapat dianalisis secara objektif dan sistematis.

Setelah perancangan sistem selesai, dilakukan pengujian dan implementasi, yang bertujuan untuk mengevaluasi sejauh mana sistem atau model yang dikembangkan dapat berfungsi dengan baik dan sesuai dengan kebutuhan penelitian. Setelah sistem diuji dan diimplementasikan, penelitian memasuki tahap analisis hasil dan pembahasan, di mana data yang telah diproses dianalisis. Sebagai tahap akhir, penelitian ini disimpulkan dengan kesimpulan dan saran.

### **3.2 Pengumpulan Data**

Tuliskan Dalam penelitian ini, data yang digunakan berasal dari data sekunder yang diperoleh melalui *Elsevier Mendeley Data Repository*, sebuah *platform online* yang berfungsi sebagai pusat penyimpanan dan berbagi data penelitian dari berbagai disiplin ilmu. Data sekunder adalah jenis data yang tidak dikumpulkan secara langsung oleh peneliti, tetapi diperoleh dari sumber lain yang telah melakukan proses pengumpulan sebelumnya. Keunggulan dari penggunaan data sekunder adalah efisiensi dalam memperoleh informasi, karena data yang telah tersedia dapat langsung diakses, dianalisis, dan diolah tanpa harus melalui tahapan pengumpulan data primer yang memerlukan waktu dan sumber daya lebih banyak.

Karena data yang digunakan berasal dari sumber sekunder, langkah pertama yang dilakukan oleh peneliti adalah pemeriksaan kualitas data untuk memastikan bahwa dataset yang digunakan layak dan valid untuk dianalisis. Terdapat beberapa aspek utama yang diperiksa, seperti keberadaan nilai yang hilang (*missing values*) dan duplikasi data, yang berpotensi menyebabkan kesalahan dalam hasil analisis jika tidak ditangani dengan baik.

Dalam penelitian ini, merupakan dataset publik yang berjudul “Dataset Ulasan Pengguna Aplikasi Telemedicine”, yang di publikasikan oleh Edi Sutoyo, dkk (2024) yang tersedia pada *Elsevier Mendeley Data Repository*. Dataset ini berisi ulasan pengguna mengenai beberapa aplikasi telemedicine terkemuka di Indonesia, seperti Alodokter dan Halodoc. Secara keseluruhan, dataset ini mencakup ribuan hingga ratusan ribu data dengan 4 atribut utama, yaitu ID atau nomor, aplikasi, review, dan label.

Dalam penelitian ini, peneliti menetapkan sebanyak 7.000 sampel data, dengan pembagian data pada Tabel 3.1.

Tabel 3.1 Detail sample data

NO	Aplikasi	Label	Jumlah Data
1.	Alodokter	Positif	1250
2.	Alodokter	Negatif	1250
3.	Alodokter	Netral	1000
4.	Halodoc	Positif	1250
5.	Halodoc	Negatif	1250
6.	Halodoc	Netral	1000
<b>Total</b>			<b>7000</b>

Hal ini dilakukan guna memastikan bahwa analisis yang dilakukan memiliki representativitas yang tinggi serta dapat mengoptimalkan proses pemrosesan dan analisis data. Pemilihan jumlah sampel ini dilakukan dengan pertimbangan bahwa

ukuran sampel yang memadai sangat berpengaruh terhadap validitas dan reliabilitas hasil penelitian, sehingga kesimpulan yang dihasilkan dapat mencerminkan persepsi masyarakat terhadap layanan kesehatan digital secara lebih akurat.

Dataset ini menjadi sumber informasi yang berharga dalam memahami bagaimana masyarakat memberikan penilaian terhadap layanan kesehatan berbasis digital. Contoh ulasan atau *review* pada dataset Tabel 3.1.

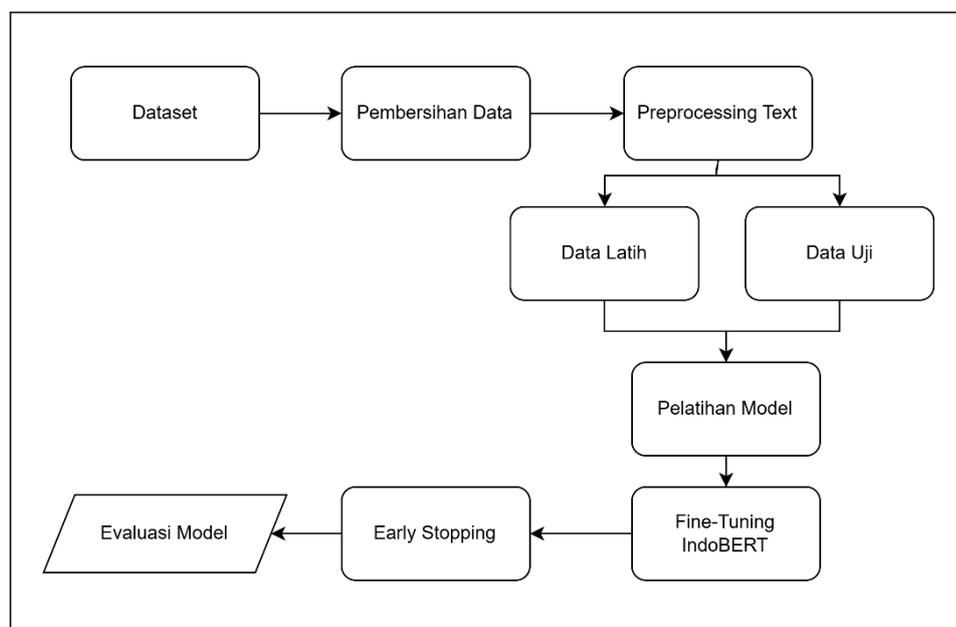
Tabel 3.2 Contoh Data Ulasan

No	Aplikasi	Review	Label
1.	Alodokter	Ingin chat dokter tapi Uda mau TF Uda pilih sesuai jasa tpi slalu nggk bisa...	Negatif
2.	Alodokter	Sayang sekali, untuk membatalkan pesanan harus WA / email admin dari alodokter terlebih dahulu.	Netral
3.	Alodokter	Saya mau berhenti berlangganan , konsultasi cuma sekali bayarnya tiap bulan baru nyadar	Negatif
4.	Alodokter	Cepat dan clear.	Positif
5.	Alodokter	Bagus walaupun sedikit lama respon nya	Positif
6.	Alodokter	Sangat membantu untuk memberikan pendidikan tentang kesehatan semoga di kedepannya bs melakukan video live streaming untuk meningkatkan diagnosis lebih lanjut...	Positif
7.	Alodokter	Kode otp tdk masuk ke sms berkali2 jd sy uninstall sj	Negatif
8.	Alodokter	Dengan aplikasi ini saya sangat terbantu untuk memeriksakan kelainan yg terjadi pada anak saya	Positif
9.	Alodokter	Aplikasi alodokter ini sangat membantu sekali	Positif
10.	Alodokter	Informasinya sangat bermanfaat makasih alodokter	Positif
11.	Alodokter	Saya Di telp barusan katanya dari cs alodokter menawarkan asuransi DBD dengan 2jta apa benar soalnya minta informasi data pribadi sampe rekening saya gk kasih takut penipuan kalo bener tolong balas yaa min...	Negatif
12.	Alodokter	Saya tidak bisa upload gambar dan video untuk gejala-gejala penyakit yang saya alami	Negatif
13.	Alodokter	Membantu sih walaupun kadang jawabanya kurang memuaskan	Netral
14.	Alodokter	Gak harus konsul langsung ke rumah skit udh bisa sdikit paham tentang kesehatan.. Tapi lebih jelas nya memang harus di konaultasikan dengan dokter di rumah sakit...	Positif
15.	Alodokter	Gak bisa koneksi... Payah	Negatif
16.	Alodokter	Saya sarankan pembayarannya harus pake dana juga. Terima kasih 🙏	Netral
17.	Alodokter	Sangat sangat bagus dan pastinya membantu masyarakat secara langsung...tingkatkan terus kualitasmu Alodokter	Positif
18.	Alodokter	Tolong hapus semua data saya. Saya memutuskan untuk tidak menggunakan aplikasi ini lagi. Terutama data rekening saya	Negatif
19.	Alodokter	Sangat menyenangkan bisa bergabung di alodokter baik2 dokternya.dan ngasih solusi arahan yg tepat. Jadi tidak bingung dan kuatir lagi deh..makasih Alodokter dan makasih dokter yg menangani saya.GBU	Positif
20.	Alodokter	Tes dulu bro.	Netral

### 3.3 Desain Sistem

Pada bagian ini, akan dijelaskan mengenai alur dan tahapan sistem yang dirancang dalam penelitian ini. Sistem yang dikembangkan bertujuan untuk melakukan analisis dan evaluasi ulasan pengguna aplikasi Alodokter dan Halodoc. Gambar 3.2 berikut menunjukkan desain sistem secara keseluruhan, yang menggambarkan alur proses dari pengambilan data hingga visualisasi hasil yang dihasilkan oleh model yang akan melalui berbagai tahapan pemrosesan data.

Setelah dilakukan dengan pengambilan dataset ulasan pengguna aplikasi Halodoc dan Alodokter yang diperoleh melalui *Elsevier Mendeley Data Repository*. Maka dilakukan proses *cleaning data*, *preprocessing*, *split data*, dan langkah-langkah selanjutnya.



Gambar 3.2 Desain Sistem

### 3.3.1 *Cleaning Data*

Pada tahap *cleaning data*, dilakukan serangkaian proses untuk memastikan bahwa data ulasan pengguna yang akan digunakan dalam analisis sentimen dan pemodelan topik berada dalam kondisi optimal. Proses ini diawali dengan pemeriksaan struktur *dataset*, termasuk jumlah kolom, jumlah data, serta tipe data dalam setiap kolom. Selanjutnya, dilakukan pengecekan terhadap missing values untuk mengidentifikasi apakah terdapat nilai kosong dalam dataset. Data yang mengandung nilai kosong dihapus untuk mencegah terjadinya kesalahan saat analisis. Selain itu, dilakukan penghapusan data duplikat guna memastikan bahwa setiap ulasan bersifat unik dan tidak memengaruhi distribusi hasil analisis secara tidak proporsional. Dengan langkah ini, dataset yang digunakan menjadi lebih bersih, terstruktur, dan siap untuk tahap pemrosesan berikutnya.

### 3.3.2 *Preprocessing*

Setelah dilakukan *cleaning data*, tahap berikutnya adalah *preprocessing*, yang bertujuan untuk memastikan kualitas data yang lebih bersih dan siap untuk dianalisis. Pada proses ini memanfaatkan *library* yang tersedia dalam bahasa pemrograman *Python*, yaitu *Pandas*, *NLTK (Natural Language Toolkit)*, dan *RE (Regular Expression)*, untuk mendeteksi kesalahan dalam dataset. *Pandas* digunakan untuk memuat dataset, memeriksa struktur data, serta menghapus baris kosong dan duplikat. Lalu *NLTK* untuk pemrosesan teks seperti tokenisasi dan *stopwords*. Sedangkan *RE (Regular Expression)* diterapkan untuk membersihkan teks dengan cara menghapus angka, karakter yang berulang, dan tanda baca yang tidak relevan. Tahapan ini dilakukan melalui lima langkah utama:

### A. *Case folding* dan *Text Cleaning*

Setelah memastikan bahwa dataset tidak mengandung data kosong atau duplikasi, dilakukan proses *cleaning teks* ulasan untuk menghilangkan elemen-elemen yang tidak relevan dalam analisis sentimen. Setiap teks dalam dataset dikonversi ke huruf kecil atau yang disebut *case folding*, guna menjaga konsistensi dalam pemrosesan bahasa alami. Kemudian, dilakukan penghapusan angka dan karakter berulang, seperti huruf yang diketik berulang kali oleh pengguna, agar teks lebih mudah dianalisis oleh model. Selain itu, tanda baca yang tidak memiliki makna dalam analisis sentimen juga dihapus, serta dilakukan penghapusan spasi berlebih untuk memastikan teks lebih rapi dan terstruktur. Hasil dari proses ini adalah teks ulasan yang lebih bersih dan siap digunakan dalam tahap tokenisasi dan analisis lebih lanjut, sehingga meningkatkan akurasi model dalam menginterpretasi opini pengguna terhadap aplikasi telemedicine. Jika semua langkah *cleaning* telah berhasil diterapkan dengan benar, dataset yang digunakan telah memenuhi standar kualitas yang diperlukan dalam penelitian. Berikut hasil setelah dilakukan *cleaning* dan *case folding* terdapat pada Tabel 3.3 :

Tabel 3.3 Proses *Cleaning* dan *Case Folding*

Sebelum <i>Cleaning</i> dan <i>Case Folding</i>	Setelah <i>Cleaning</i> dan <i>Case Folding</i>
saya sangat puasssssss 🤝 🤝 🤝	saya sangat puas
Dokter nya menjelaskan dengan bagus dan bisa dipahami, baguss bangettt, saya selalu konsultasi ke alo dokter tentang Kesehatan	dokter nya menjelaskan dengan bagus dan bisa dipahami baguss banget saya selalu konsultasi ke alo dokter tentang Kesehatan
Ngga ada respon dari dokter setelah mengisi bagian pertanyaan :)	ngga ada respon dari dokter setelah mengisi bagian pertanyaan

## **B. Normalisasi**

Tahap normalisasi dalam *preprocessing* bertujuan untuk mengubah kata-kata tidak baku yang sering digunakan dalam teks ulasan menjadi bentuk yang lebih standar atau baku, sesuai dengan kaidah bahasa yang benar. Hal ini penting karena banyaknya variasi penggunaan bahasa dalam ulasan, seperti singkatan, bahasa gaul, atau kata-kata tidak baku, yang dapat mengurangi efektivitas analisis sentimen jika tidak dinormalisasi terlebih dahulu. Misalnya, kata-kata seperti “gpp” yang berarti “tidak apa-apa” atau “ga” yang berarti “tidak” akan diubah menjadi bentuk yang lebih baku sesuai dengan kamus normalisasi yang telah ditentukan. Dengan demikian, proses ini memastikan bahwa teks yang dianalisis lebih konsisten dan dapat dipahami dengan lebih baik oleh model.

Selain itu, normalisasi ini juga membantu dalam mengurangi keragaman bahasa yang tidak perlu, yang pada akhirnya dapat meningkatkan akurasi model dalam melakukan klasifikasi sentimen dan pemodelan topik. Beberapa contoh lain dari kata-kata yang dinormalisasi termasuk “emg” menjadi “memang” atau “krn” menjadi “karena”. Dengan cara ini, teks ulasan yang awalnya mungkin mengandung banyak variasi bahasa gaul atau singkatan yang tidak dikenali oleh model, kini lebih terstruktur dan dapat diproses lebih efisien. Normalisasi juga memastikan bahwa kata-kata dengan arti yang sama, meskipun dieja dengan cara yang berbeda, diperlakukan secara konsisten, sehingga hasil analisis lebih representatif dan dapat diandalkan. Berikut hasil setelah dilakukan normalisasi pada Tabel 3.4 :

Tabel 3.4 Proses Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
sangat membantu apalagi kalau orang yg ga sempet ke dokter langsung	sangat membantu apalagi kalau orang yang tidak sempet ke dokter langsung
dua hari lalu saya abis konsultasi di alodokter alhamdulillah skrg saya udh sembuh karna dikasih resep untuk obatnya terimakasih alodokter	dua hari lalu saya abis konsultasi di alodokter alhamdulillah sekarang saya sudah sembuh karna dikasih resep untuk obatnya terimakasih alodokter
sangat sangat bermanfaat dokternya responnya cepat ga sampe menit sudah ada dokter yg merespon dokternya baik baik dan aplikasinya recommended	sangat sangat bermanfaat dokternya responnya cepat tidak sampe menit sudah ada dokter yang merespon dokternya baik baik dan aplikasinya recommended

### C. Tokenisasi

Pada tahap tokenisasi (*tokenizing*) dalam proses *preprocessing* teks, tujuan utama adalah memecah teks yang telah dinormalisasi menjadi unit-unit yang lebih kecil yang disebut token. Token ini biasanya berupa kata-kata individual, namun dapat juga berupa tanda baca atau simbol lainnya tergantung pada implementasi yang digunakan. Proses ini sangat penting untuk mempersiapkan data agar bisa diproses lebih lanjut oleh model analisis sentimen atau pemodelan topik. Misalnya, dalam ulasan pengguna aplikasi telemedicine, kata-kata seperti “sangat baik” atau “tidak efektif” akan dipisah menjadi token “sangat”, “baik”, “tidak”, “efektif”. Proses tokenisasi ini juga membantu dalam mengurangi ketergantungan pada konteks kalimat, karena setiap token yang terpisah dapat diproses secara individual. Dalam hal ini, tokenization memainkan peran penting dalam mempersiapkan data yang lebih bersih dan lebih mudah dianalisis oleh model berbasis *natural language processing* (NLP), yang selanjutnya akan mengklasifikasikan sentimen atau mengidentifikasi topik-topik utama yang dibahas dalam ulasan pengguna. Berikut hasil setelah dilakukan *tokenizing* (Tabel 3.5):

Tabel 3.5 Proses *Tokenizing*

<b>Sebelum <i>Tokenizing</i></b>	<b>Setelah <i>Tokenizing</i></b>
namanya online tidak terlalu leluasa dalam bertanya namun informasi yang diberikan dengan nada ramah dan cukup lengkap bisa membantu sekali	['namanya', 'online', 'leluasa', 'informasi', 'nada', 'ramah', 'lengkap', 'membantu']
banyak penipuan nya padahal aplikasi bagus dan sangat membantu tetapi pengguna benar benar merasa sangat terganggu karena sering dapat telepon mengatas namakan alodokter keamanan nya mohon di tingkatkan	['banyak', 'penipuan', 'nya', 'padahal', 'aplikasi', 'bagus', 'dan', 'sangat', 'membantu', 'tetapi', 'pengguna', 'benar', 'benar', 'merasa', 'sangat', 'terganggu', 'karena', 'sering', 'dapat', 'telepon', 'mengatas', 'namakan', 'alodokter', 'keamanan', 'nya', 'mohon', 'di', 'tingkatkan']
respon pak bu dokter sangat cepat dalam menjawabnya yang lelet internet gw dan pastinya konsultasi free	['respon', 'pak', 'bu', 'dokter', 'sangat', 'cepat', 'dalam', 'menjawabnya', 'yang', 'lelet', 'internet', 'gw', 'dan', 'pastinya', 'konsultasi', 'free']

#### ***D. Stopwords Removal***

Pada tahap *Stopwords Removal* dalam proses *preprocessing*, tujuan utamanya adalah menghapus kata-kata yang sering muncul dalam teks tetapi tidak memberikan informasi signifikan untuk analisis sentimen atau pemodelan topik. *Stopwords* biasanya mencakup kata-kata fungsional dalam bahasa seperti “dan”, “yang”, “di”, serta kata-kata yang bersifat umum dan tidak relevan dengan konteks analisis. Dalam penelitian ini, daftar stopwords yang digunakan mencakup kata-kata umum dalam bahasa Indonesia, seperti “sih”, “saya”, “gak”, “ya”, dan lainnya yang sering muncul dalam ulasan pengguna namun tidak menyumbang banyak makna dalam analisis. Proses ini penting karena dengan menghapus *stopwords*, model dapat lebih fokus pada kata-kata yang benar-benar menggambarkan opini pengguna tentang aplikasi telemedicine, sehingga meningkatkan akurasi dan relevansi hasil analisis. Hasil dari tahap ini adalah teks yang lebih bersih dan

terfokus, di mana hanya kata-kata yang relevan yang dipertahankan untuk langkah selanjutnya dalam pemrosesan.

### 3.3.3 Pelatihan Model

Pada tahap pelatihan model, langkah pertama yang dilakukan adalah mengonversi label sentimen dalam dataset, yaitu “negatif”, “positif”, dan “netral”, diubah menjadi bentuk numerik untuk memudahkan pemrosesan oleh model, di mana “negatif” dikodekan sebagai 0, “positif” sebagai 1, dan “netral” sebagai 2. Proses ini penting karena model IndoBERT membutuhkan input dalam bentuk angka untuk memprosesnya. Selanjutnya, tokenisasi dilakukan menggunakan tokenizer IndoBERT, yang merupakan tahap yang berbeda dari tokenisasi yang dilakukan pada tahap *preprocessing*. Pada *preprocessing*, tokenisasi dilakukan secara sederhana dengan membagi teks menjadi kata-kata individual, sedangkan pada pelatihan model, tokenisasi IndoBERT memanfaatkan model berbasis BERT yang lebih kompleks untuk mengonversi teks menjadi token yang sesuai dengan format *input* yang diterima oleh model BERT. Tokenisasi ini melibatkan pemecahan teks menjadi unit-unit yang lebih kecil, atau token, yang memungkinkan model untuk menangkap konteks dan hubungan antar kata dalam kalimat dengan lebih baik, serta memastikan bahwa setiap token terhubung dengan representasi yang lebih mendalam. Teks ulasan kemudian diubah menjadi representasi token dengan panjang maksimum 128 kata, sesuai dengan standar *input* model BERT.

Setelah proses tokenisasi selesai, data pelatihan dan validasi dimasukkan ke dalam *dataset PyTorch* yang menghubungkan tokenisasi dengan label numerik

yang telah diubah sebelumnya. Data ini kemudian diproses menggunakan *DataLoader*, yang membagi dataset menjadi *batch-batch* kecil dengan *batch size* 16 guna efisiensi selama pelatihan dan validasi. Proses ini dilakukan agar mengurangi potensi *overfitting* dengan memvalidasi kinerja model menggunakan data yang tidak dilatih sebelumnya.

### 3.3.4 *Fine-Tuning* IndoBERT

Pada tahap *Fine-Tuning* IndoBERT, model dilatih untuk mempelajari dan menyesuaikan parameter internalnya berdasarkan data pelatihan yang telah diproses. Pada proses ini menggunakan *library BertTokenizer* yang terdapat pada *transformers* model BERT yang digunakan untuk untuk mengonversi teks menjadi token yang dapat diproses oleh model BERT dan *BertForSequenceClassification* yang digunakan untuk membangun model klasifikasi berdasarkan model BERT, yang dipakai dalam tahap training untuk mengklasifikasikan ulasan menjadi sentimen. Kemudian juga digunakan algoritma AdamW, yang merupakan varian dari *Adam optimizer* yang menggabungkan *weight decay* untuk regularisasi. Rumus untuk momentum pertama dan momentum kedua dalam AdamW (*Adaptive Moment Estimation with Weight Decay*) adalah algoritma optimisasi yang digunakan untuk memperbarui parameter model dalam pelatihan deep learning, seperti *fine-tuning* model IndoBERT. Algoritma ini merupakan varian dari Adam, yang mengkombinasikan dua jenis momentum (momentum pertama dan momentum kedua) untuk meningkatkan kecepatan konvergensi dan stabilitas selama pelatihan. AdamW mengatasi masalah *overfitting* dengan menambahkan *weight decay*, yang bertujuan untuk mengurangi ukuran parameter model dan

mencegah model menjadi terlalu kompleks. Proses pembaruan parameter dalam AdamW dilakukan melalui tiga langkah utama: pembaruan momentum pertama (estimasi gradien), pembaruan momentum kedua (estimasi kuadrat gradien), dan pembaruan parameter itu sendiri. Agar model dapat bekerja secara optimal, proses *fine-tuning* dilakukan dengan parameter Tabel 3.6:

Tabel 3.6 Parameter *Fine-Tuning* IndoBERT

<i>Hyperparameter</i>	Nilai
<i>Learning Rate</i>	3e-5
$\beta_1$	0.9
$\beta_2$	0.999
$\epsilon$	1e-8
<i>Weight Decay</i>	0.01

Proses pertama yang dilakukan adalah menghitung momentum pertama ( $m_t$ ) dan momentum kedua ( $v_t$ ) untuk setiap parameter model berdasarkan gradien yang dihitung selama *backpropagation*. Misalnya pada iterasi pertama ( $t = 1$ ), momentum pertama dan momentum kedua diinisialisasi menjadi 0, dan pembaruan dilakukan berdasarkan gradien pertama yang dihitung. Berikut adalah contoh perhitungan untuk iterasi pertama dengan  $gt = 0.5$  yang dilakukan berdasarkan persamaan 2.1 dan 2.2:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot gt$$

$$m_1 = 0.9 \cdot 0 + (1 - 0.9) \cdot 0.5 = 0.05$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot gt^2$$

$$v_1 = 0.999 \cdot 0 + (1 - 0.999) \cdot 0.5^2 = 0.0005$$

Selanjutnya, koreksi bias diterapkan untuk memperbaiki estimasi momentum pada awal pelatihan, karena pada iterasi pertama  $m_t$  dan  $v_t$  memiliki

nilai kecil yang terdistorsi. Koreksi bias dihitung dengan rumus pada persamaan 2.3 dan 2.4 sebagai berikut:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} = \frac{0.05}{1 - 0.9} = 0.5$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} = \frac{0.0005}{1 - 0.999} = 0.5$$

Dengan bias correction, estimasi momentum menjadi lebih akurat, terutama pada iterasi awal. Setelah itu, pembaruan parameter model dilakukan dengan rumus AdamW, dimana parameter model pada iterasi ke-0 adalah  $\theta_0=1$ . Maka pembaruan parameter dihitung berdasarkan persamaan 2.5 sebagai berikut:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta \cdot \lambda \cdot \theta_t$$

$$\theta_1 = 1 - 2e^{-5} \cdot \frac{0.5}{\sqrt{0.5_t + 1e^{-8}}} - 2e^{-5} \cdot 0.01 \cdot 1$$

$$\theta_1 = 1 - 2e^{-5} \cdot \frac{0.5}{0.7071 + 1e^{-8}} - 2e^{-5} \cdot 0.01$$

$$\theta_1 = 1 - 2e^{-5} \cdot 0.7071 - 2e^{-7}$$

$$\theta_1 = 1 - 0.00001414 + 0.00000002$$

$$\theta_1 = 0.99998584$$

Hasil  $\theta_1=0.99998584$  menunjukkan bahwa parameter model telah diperbarui sebanyak 0.00001416. Sedikit penurunan ini mencerminkan penerapan weight decay yang benar dalam algoritma AdamW, yang membantu mencegah model menjadi terlalu kompleks dan *overfitting* terhadap data pelatihan..

Pembaruan parameter ini memungkinkan model IndoBERT untuk membuat prediksi yang lebih baik pada data pelatihan. Proses ini dilakukan secara bertahap, di mana model terus belajar dan menyesuaikan diri pada setiap langkah untuk meningkatkan akurasi dan kinerja model secara keseluruhan. Dengan rumus ini, AdamW berhasil memperbarui parameter model secara efisien dan stabil, meminimalkan *loss*, dan meningkatkan kemampuan model untuk menggeneralisasi ke *data testing* yang tidak terlihat sebelumnya.

Langkah selanjutnya adalah proses *forward pass*, di mana input berupa teks yang telah melalui tokenisasi diubah menjadi representasi numerik (*logits*). *Logits* ini kemudian digunakan untuk menghitung probabilitas dengan menggunakan fungsi *softmax*. Fungsi *softmax* berfungsi untuk mengubah skor mentah menjadi probabilitas kelas. Berikut contoh perhitungan *softmax* pada ulasan “Sangat membantu sy utk menambah pengetahuan kesehatan dan pencegahan penyakit dg pola hidup sehat” dengan rumus berdasarkan persamaan 3.7 didapatkan hasil pada Tabel 3.7.

Tabel 3.7 Nilai Logits

No	Kelas	Nilai
1.	Negatif	-3.2216
2.	Positif	2.3646
3.	Netral	1.1295

#### 1. Hitung Eksponensial dari Setiap Logit

$$e^{z_{Negatif}} = e^{-3.2216} = 0.0399$$

$$e^{z_{Positif}} = e^{2.3646} = 10.6379$$

$$e^{z_{netral}} = e^{1.1295} = 3.0941$$

2. Jumlahkan Semua Eksponensial

$$\sum_j e^{z_j} = 0.0399 + 10.6379 + 3.0941 = 13.7719$$

3. Hitung Probabilitas untuk Setiap Kelas: Menggunakan rumus *softmax*

$$e^{z_{negatif}} = \frac{0.0399}{13.7719} = 0.0029$$

$$e^{z_{positif}} = \frac{10.6379}{13.7719} = 0.7723$$

$$e^{z_{netral}} = \frac{3.0941}{13.7719} = 0.2247$$

Berdasarkan hasil ini, model memprediksi bahwa ulasan tersebut memiliki sentimen positif dengan probabilitas 77.23%, sementara kemungkinan untuk negatif dan netral sangat rendah (masing-masing 0.29% dan 22.47%). Ini menunjukkan bahwa model menilai ulasan ini lebih bersifat positif.

Selanjutnya, untuk mengukur ketidaksesuaian antara probabilitas prediksi model dan label sentimen yang sebenarnya, digunakan *cross-entropy loss*. Model yang baik akan menghasilkan *loss* yang semakin kecil. Berdasarkan perhitungan *softmax* yang dilakukan sebelumnya, berikut perhitungan *Cross-Entropy Loss* berdasarkan persamaan 2.7:

$$\mathcal{L} = - \sum_{c=1}^c y_c \log(p_c)$$

$$L = -(y_{negatif} \log(p_{negatif}) + y_{netral} \log(p_{netral}) + y_{positif} \log(p_{positif}))$$

$$L = -(0 \cdot \log(0.0029) + 1 \cdot \log(0.2247) + 0 \cdot \log(0.7723))$$

$$L = -\log(0.2247)$$

$$L = -(-0.2583) = 0.2583$$

Hasil perhitungan *cross-entropy loss* yang dihitung untuk ulasan tersebut, menunjukkan bahwa nilai *loss* yang diperoleh adalah 0.2583 setara dengan 25.83%, yang artinya model telah memberikan prediksi dengan akurasi cukup baik. Dalam hal ini, model memprediksi kelas Positif dengan probabilitas 77.23%, yang sesuai dengan label asli ulasan tersebut, dan menghasilkan *loss* yang rendah, yang menunjukkan perbedaan kecil antara probabilitas prediksi dan label yang sebenarnya.

Setelah dilakukan keseluruhan proses *fine-tuning* tersebut, maka diterapkan teknik *early stopping* untuk menghindari *overfitting* dan memastikan model tidak terlalu terfokus pada data pelatihan. Proses ini melibatkan pemantauan *validation accuracy* setiap kali selesai satu *epoch*. Apabila *validation loss* terus mengalami penurunan, pelatihan akan dilanjutkan. Namun, jika tidak ada perbaikan dalam *validation accuracy* berdasarkan *patience* yang ditentukan (misalnya 3 *epoch*), maka pelatihan akan dihentikan lebih awal. Tujuan dari hal ini adalah untuk menghindari pembentukan model yang terlalu rumit dan tidak dapat menggeneralisasi dengan baik pada data baru. Sebuah *counter* digunakan untuk menghitung jumlah *epoch* yang tidak menunjukkan perbaikan pada *validation accuracy*, dan apabila nilai *counter* mencapai ambang batas yang ditentukan, pelatihan dihentikan. Model yang memiliki *best validation loss* selama pelatihan akan disimpan sebagai model final.

### 3.3.5 Evaluasi Model

Evaluasi Model dalam penelitian ini bertujuan untuk mengukur kinerja *fine-tuning* IndoBERT dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi telemedicine. Metode evaluasi yang digunakan dalam penelitian ini adalah *Confusion Matrix* serta metrik akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi ini penting untuk menilai sejauh mana model mampu membedakan sentimen positif, netral, dan negatif secara akurat seperti penelitian yang dilakukan oleh (Mahardhika dkk., 2015).

*Confusion Matrix* digunakan sebagai alat untuk menganalisis distribusi prediksi model terhadap data yang sebenarnya. Matriks ini memberikan gambaran tentang jumlah prediksi yang benar dan salah dalam setiap kelas sentimen. Struktur dasar *Confusion Matrix* 3x3 ditampilkan dalam Tabel 3.8 berikut:

Tabel 3.8 *Confusion Matrix* dalam analisis sentimen

Aktual	Prediksi		
	Positif	Netral	Negatif
Positif	TP	FN	FN
Netral	FP	TN	TN
Negatif	FP	TN	TN

Keterangan:

TP : *True Positive* (Jumlah data positif dan benar positif)

TN : *True Negative* (Jumlah data negatif dan benar negatif)

FP : *False Positive* (Jumlah data positif yang sebenarnya negatif)

FN : *False Negative* (Jumlah data negatif yang sebenarnya positif)

Untuk mengevaluasi model, dilakukan perhitungan dengan metrik akurasi, *precision*, *recall*, dan *F1-score*. Perhitungan dilakukan berdasarkan rumus yang dinyatakan sebagai berikut (Wabang dkk., 2022):

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP \text{ Total}}{\text{Dataset Total}} \quad (3.1)$$

$$Presisi = \frac{TP}{TP+FP} = \frac{TP}{\text{Prediksi Total}} \quad (3.2)$$

$$Recall = \frac{TP}{TP+FN} = \frac{TP}{Aktual\ Total} \quad (3.3)$$

$$F1 - score = 2 \times \frac{presisi \times recall}{presisi + recall} \quad (3.4)$$

Evaluasi sistem ini bertujuan untuk memastikan bahwa model tidak hanya akurat tetapi juga konsisten dalam menangkap pola sentimen yang ada dalam data. Akurasi, *precision*, *recall*, dan *F1-score* adalah metrik evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Akurasi digunakan untuk mengukur persentase prediksi yang benar dari keseluruhan data, sedangkan *precision* menunjukkan seberapa tepat model dalam memprediksi kelas positif. *Recall* mengukur sejauh mana model dapat menangkap seluruh instance positif yang sebenarnya ada, sementara *F1-score* merupakan keseimbangan antara *precision* dan *recall*, yang penting dalam dataset yang tidak seimbang. Keempat metrik ini digunakan bersama untuk memberikan evaluasi yang lebih akurat dan adil terhadap performa model. Jika nilai akurasi, *precision*, *recall*, dan *F1-score* tinggi untuk semua kelas sentimen, maka model dapat dikatakan bekerja dengan baik dalam klasifikasi sentimen.

### 3.4 Skenario Pengujian

Pada penelitian ini, akan dilakukan 2 skenario pengujian dengan variasi data split dan jumlah *epoch*. Eksperimen pertama akan menggunakan teknik *Grid Search* untuk mengoptimalkan *hyperparameter learning rate*. Dalam pengujian ini, berbagai kombinasi nilai *learning rate* akan diuji secara otomatis menggunakan *Grid Search*, yang memungkinkan pencarian *hyperparameter* terbaik secara efisien dalam ruang yang telah ditentukan. Nilai *learning rate* yang diuji akan berkisar dari

1e-5 hingga 5e-5. Dengan memanfaatkan Grid Search, eksperimen ini akan menguji berbagai nilai *learning rate* secara bersamaan dan memilih kombinasi terbaik yang menghasilkan kinerja model optimal. Scenario ini dilakukan dengan variasi split data 80:20.

Skenario kedua, di sisi lain, akan menggunakan jumlah *epoch* yang fleksibel dan *early stopping* untuk menghindari *overfitting*. Dalam eksperimen ini, pelatihan model akan dihentikan jika *validation accuracy* tidak menunjukkan perbaikan setelah beberapa *epoch* berturut-turut. *Early stopping* diharapkan dapat menghasilkan model yang tepat dengan lebih efisien serta mampu menggeneralisasi dengan lebih baik. Dalam eksperimen ini akan dilakukan dengan tiga variasi split data yang dibagi menjadi data training:data testing. Yang dilakukan dalam tiga cara, yaitu 90:10, 80:20, dan 70:30. Hal ini dilakukan untuk mengukur pengaruh split data terhadap performa model serta mengevaluasi bagaimana *epoch* fleksibel dan *early stopping* mempengaruhi kinerja model pada dataset yang dibagi dengan proporsi yang berbeda. Skenario pengujian dapat dilihat pada tabel 3.9.

Tabel 3.9 Skenario Pengujian

Skenario	Deskripsi	Data Training	Data Testing
Skenario 1	Menggunakan teknik <i>Grid Search</i> untuk menentukan dan mengoptimalkan <i>hyperparameter learning rate</i> dengan menetapkan 2 <i>epoch</i>	80%	20%
Skenario 2	Menggunakan jumlah <i>epoch</i> fleksibel dengan <i>early stopping</i> untuk menghindari <i>overfitting</i>	90%	10 %
		80%	20%
		70%	30%

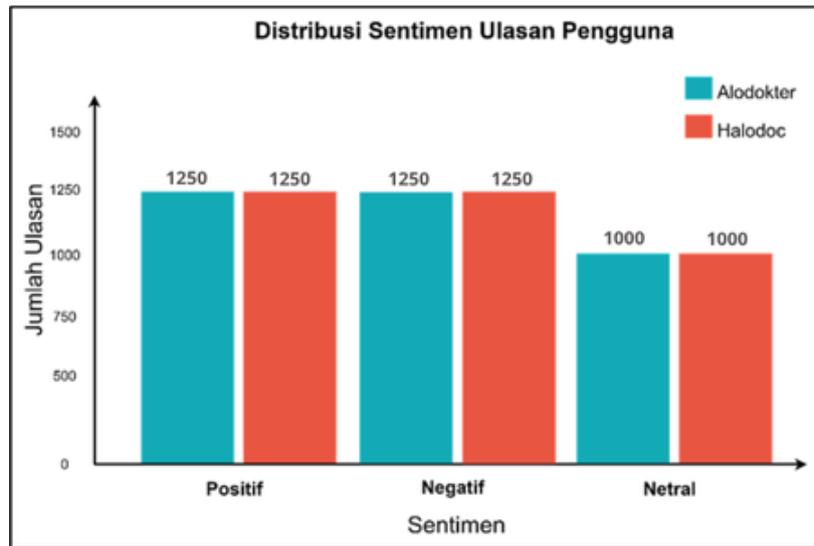
## **BAB IV**

### **UJI COBA DAN PEMBAHASAN**

#### **4.1 Pengumpulan Data**

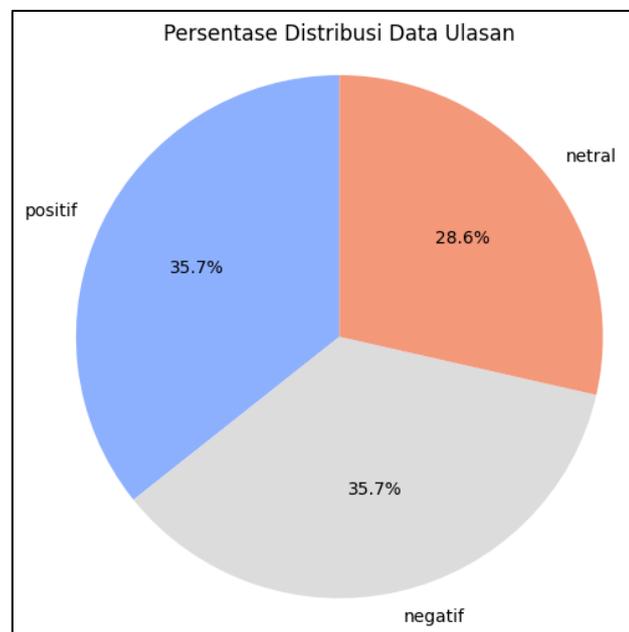
Data penelitian yang digunakan dalam studi ini merupakan data sekunder yang diperoleh dari repositori publik *Elsevier Mendeley Data Repository*. Dataset yang digunakan berjudul “Dataset Ulasan Pengguna Aplikasi Telemedicine” yang dipublikasikan oleh Edi Sutoyo, dkk. pada tahun 2024. Dataset yang digunakan dari dua aplikasi telemedicine populer di Indonesia, yaitu Halodoc dan Alodokter.

Dataset ini terdiri atas tiga label sentimen yaitu positif, negatif, dan netral, yang merepresentasikan persepsi pengguna terhadap pengalaman mereka saat menggunakan aplikasi. Dataset ini telah difilter dan disesuaikan sehingga terdiri dari total 7.000 ulasan, dengan distribusi merata di antara aplikasi dan label sentimen. Secara rinci, setiap aplikasi masing-masing memiliki 3.500 ulasan, terdiri atas 1.250 ulasan positif, 1.250 ulasan negatif, dan 1.000 ulasan netral. Sebagaimana perbandingan data dapat dilihat pada Gambar 4.1.



Gambar 4.1 Distribusi Sentimen Ulasan Pengguna

Visualisasi presentase distribusi data ulasan antara positif, negatif, dan netral ditunjukkan pada gambar 4.2



Gambar 4.2 Distribusi Data Ulasan

Gambar 4.2 menunjukkan persentase data yang terbagi dalam tiga kategori: positif (35,7%), negatif (35,7%), dan netral (28,6%) atau sama dengan . Pelabelan tersebut telah tercantum pada dataset yang didapat pada *Elsevier Mendeley Data Repository* yang dipublikasikan oleh Edi Sutoyo, dkk yang dapat dilihat pada Tabel 4.1

Tabel 4.1 Contoh Dataset

No	Aplikasi	Review	Label
1.	halodoc	Mantap	positif
2.	alodokter	Saya baru downlod semoga bisa bantu masalah saya ya .	netral
3.	alodokter	Trimakasih dokter sdh sangat membantu konsultasi saya... :D	positif
4.	halodoc	Belum bisa merasakan kegunaannya	netral
5.	halodoc	tidak bisa bayar di tempat	negatif
6.	alodokter	Bagus sekali	positif
7.	halodoc	Mungkin dokternya sibuk, agak lama balasnya	negatif
8.	alodokter	Mudah mudahan aplikasinya bagus	netral
9.	alodokter	Oke,saya cobak dulu ya	netral
10.	halodoc	Hallo doc perbaharui lagi masa e-wallet gabisa di tarik tunai	negatif

## 4.2 Uji Coba

Proses uji coba dalam penelitian ini dilaksanakan menggunakan platform Google Colaboratory, dengan tujuan utama untuk menilai kinerja model *fine-tuning* IndoBERT dalam menganalisis sentimen dari ulasan pengguna aplikasi telemedicine, yaitu Halodoc dan Alodokter. Eksperimen yang dilakukan bertujuan untuk mengukur efektivitas model dalam mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral. Uji coba ini melibatkan variasi parameter, seperti *split data*, jumlah *epoch*, dan optimasi *hyperparameter* pada *learning rate*. *Grid Search* diterapkan untuk mengoptimalkan nilai *learning rate*, dan *early stopping* digunakan untuk mencegah terjadinya *overfitting*. Sementara Evaluasi dilakukan dengan menggunakan *confusion matrix*. Semua tahapan ini dijalankan

menggunakan bahasa pemrograman *Python* untuk memastikan efisiensi dan konsistensi dalam pelatihan dan evaluasi model.

#### 4.2.1 *Cleaning Data*

Pada tahap *cleaning data*, dua proses dilakukan untuk memastikan dataset ulasan pengguna berada dalam kondisi optimal untuk analisis sentimen. Pada tahap ini dataset akan diperiksa menggunakan *library Pandas*. Kemudian dilakukan pengecekan terhadap *missing values* dengan menggunakan fungsi *.isnull().sum()* untuk memastikan tidak ada data yang hilang. Jika terdapat nilai kosong, baris dengan *missing values* dihapus menggunakan method *.dropna()*. Jika terdapat nilai kosong, baris dengan *missing values* dihapus. Selain itu, data duplikat yang dapat mempengaruhi hasil analisis juga dihapus dengan menggunakan metode *.drop\_duplicates()*. Proses ini dilakukan untuk memastikan bahwa setiap ulasan dalam dataset unik dan tidak ada data yang terduplikasi. Hasil dari implementasi *cleaning data* dapat dilihat pada Gambar 4.3.

```

♦ Data Setelah Dibersihkan:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7000 entries, 0 to 6999
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   no           7000 non-null   int64
1   aplikasi    7000 non-null   object
2   review      7000 non-null   object
3   label       7000 non-null   object
dtypes: int64(1), object(3)
memory usage: 218.9+ KB
None

 Dataset telah dibersihkan dan disimpan

```

Gambar 4.3 Hasil Cleaning Data

Dataset yang berjumlah 7.000 entri ini terdiri dari empat kolom utama, yaitu nomor urut, aplikasi, *review*, dan label. Setelah dilakukan pembersihan, setiap kolom telah terbebas dari nilai kosong (*missing values*) dan duplikasi, yang memastikan bahwa data yang digunakan adalah akurat dan tidak terdistorsi. Hal ini tercermin dari hasil yang menunjukkan bahwa semua kolom memiliki 7.000 entri non-null, dengan tipe data yang sesuai untuk analisis lebih lanjut.

#### 4.2.2 *Preprocessing*

Pada tahap *preprocessing*, beberapa langkah penting dilakukan untuk mempersiapkan data ulasan pengguna agar dapat diproses lebih lanjut dalam analisis sentimen. Langkah-langkah ini memastikan bahwa data yang digunakan lebih bersih, terstruktur, dan siap untuk dianalisis secara efektif oleh model. Berikut adalah penjelasan rinci dari setiap langkah yang dilakukan dalam *preprocessing*:

a. *Case Folding* dan *Text Cleaning*

Langkah pertama dalam *preprocessing* adalah *case folding*, yang bertujuan untuk mengubah seluruh teks ulasan menjadi huruf kecil. Proses ini dilakukan untuk menghilangkan perbedaan antara huruf kapital dan huruf kecil yang tidak relevan dalam analisis sentimen. Dengan mengubah seluruh teks menjadi huruf kecil, model dapat memperlakukan kata-kata seperti “Baik” dan “baik” sebagai kata yang sama, sehingga mengurangi variabilitas yang tidak perlu. Fungsi *.lower()* dalam *Python* digunakan untuk mengonversi seluruh teks menjadi huruf kecil, memastikan konsistensi dalam pemrosesan teks.

Setelah *case folding*, tahap berikutnya adalah pembersihan teks (*text cleaning*). Proses ini menggunakan *regular expressions* (re) yang diterapkan

dengan metode *.re.sub()* untuk menghapus angka, menghilangkan huruf yang berulang (misalnya, “loooove” menjadi “love”), dan mengganti tanda baca dengan spasi. Tanda baca yang tidak relevan, seperti tanda seru atau koma, dapat mengganggu proses analisis sentimen, sehingga dihapus untuk menyederhanakan teks. Selain itu, spasi berlebih juga dihapus untuk memastikan teks yang lebih rapi dan terstruktur. Dengan langkah ini, teks yang digunakan dalam analisis menjadi lebih bersih dan lebih mudah diproses lebih lanjut. Hasil dari *case folding* dan *text cleaning* dapat dilihat pada tabel 4.2.

Tabel 4.2 Hasil *Case Folding* dan *Text Cleaning*

Sebelum <i>Case Folding</i> dan <i>Text Cleaning</i>	Setelah <i>Case Folding</i> dan <i>Text Cleaning</i>
Apaaa <sup>2</sup> an ini saya tiba <sup>2</sup> di telfon di tawari Asuransi dgn potongan 39rb/BLN ... Tolong dari pihak APP Alodokter jgn ganggu privasi kami dgn menelfon satu <sup>2</sup> no kami yg terdaftar di aplikasi...	apa <sup>2</sup> an ini saya tiba di telfon di tawari asuransi dgn potongan rb bln tolong dari pihak app alodokter jgn ganggu privasi kami dgn menelfon satu <sup>2</sup> no kami yg terdaftar di aplikasi

#### b. Normalisasi

Setelah teks dibersihkan, tahap normalisasi dilakukan untuk mengganti kata-kata tidak baku atau singkatan dengan kata baku yang sesuai. Dalam dataset ulasan pengguna, sering ditemukan kata-kata singkatan atau bahasa gaul yang tidak akan dipahami dengan baik oleh model jika dibiarkan begitu saja. Misalnya, kata “gpp” yang berarti “tidak apa-apa” atau “gak” yang berarti “tidak” perlu digantikan dengan bentuk baku yang sesuai untuk meningkatkan konsistensi teks. Proses normalisasi ini menggunakan kamus yang telah disusun, dan diterapkan dalam fungsi *.normalize\_text* untuk menggantikan kata-kata tersebut. Dengan normalisasi, variasi kata yang tidak standar akan diseragamkan menjadi bentuk yang lebih baku,

sehingga model dapat memahami makna secara lebih efektif. Hasil dari normalisasi teks dapat dilihat pada tabel 4.3.

Tabel 4.3 Hasil Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
Aplikasi ini sangat bagus dapat membantu memberikan pertanyaan yg sy harapkan tentang kesehatan	aplikasi ini sangat bagus dapat membantu memberikan pertanyaan yang saya harapkan tentang kesehatan

### c. Tokenisasi

Setelah teks di normalisasi, tahap berikutnya adalah tokenisasi, di mana teks dibagi menjadi unit-unit yang lebih kecil yang disebut token. Tokenisasi memungkinkan model untuk menganalisis teks dalam bentuk kata-kata yang terpisah, bukan dalam bentuk kalimat utuh. Proses ini dilakukan dengan menggunakan fungsi *word\_tokenize* dari NLTK yang membagi teks menjadi kata-kata terpisah. Dengan tokenisasi, kalimat seperti “Aplikasi ini sangat membantu” akan dipisahkan menjadi token [“Aplikasi”, “ini”, “sangat”, “membantu”], yang memudahkan model dalam mengenali pola dan makna dari kata-kata tersebut. Hasil dari tokenisasi teks dapat dilihat pada tabel 4.4.

Tabel 4.4 Hasil Tokenisasi

Sebelum Tokenisasi	Setelah Tokenisasi
mudah menggunakannya klo bisa banyakin rekan yang sudah kerjasama di hallodoc sehingga untuk pengiriman bisa disesuaikan dengan lokasi pemesanan kasian drivernya klo jauh ngirim thx	['mudah', 'menggunakannya', 'klo', 'bisa', 'banyakin', 'rekan', 'yang', 'sudah', 'kerjasama', 'di', 'hallodoc', 'sehingga', 'untuk', 'pengiriman', 'bisa', 'dिसesuaikan', 'dengan', 'lokasi', 'pemesanan', 'kasian', 'drivernya', 'klo', 'jauh', 'ngirim', 'thx']

### d. Stopword Removal

Langkah terakhir dalam *preprocessing* adalah *stopword removal*, yang bertujuan untuk menghapus kata-kata umum yang sering muncul dalam teks namun tidak memberikan kontribusi signifikan terhadap analisis sentimen. Kata-kata

seperti “dan”, “yang”, “di”, serta kata ganti dan kata sambung tidak relevan untuk analisis karena tidak membantu dalam pengidentifikasian sentimen. Pada proses ini, NLTK digunakan untuk mengakses daftar stopwords dalam bahasa Indonesia, yang mencakup kata-kata umum yang sering ditemukan dalam teks. Selain itu, daftar *stopwords* kustom juga dibuat untuk mencakup kata-kata yang lebih spesifik, seperti “sih”, “ya”, atau “dong”. Dengan menghapus kata-kata ini, model dapat fokus pada kata-kata yang lebih relevan untuk analisis sentimen, meningkatkan akurasi dalam memprediksi sentimen yang terkandung dalam teks. Hasil dari normalisasi teks dapat dilihat pada tabel 4.5.

Tabel 4.5 Hasil *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
['tolong', 'dong', 'jangan', 'begitu', 'dok', 'tidak', 'ada', 'konsul', 'chat', 'apa', 'apa', 'masa', 'masih', 'tetep', 'kepotong', 'debit', 'otomatis']	['tolong', 'dok', 'konsul', 'chat', 'tetep', 'kepotong', 'debit', 'otomatis']

Dengan langkah-langkah *preprocessing* ini, dataset yang digunakan dalam penelitian ini dipersiapkan secara sistematis untuk analisis lebih lanjut. Teks yang telah dibersihkan, dinormalisasi, dan diproses melalui tokenisasi dan penghapusan *stopwords* memastikan bahwa model dapat memproses data dengan lebih efisien, menghasilkan hasil yang lebih akurat dan dapat diandalkan dalam analisis sentimen ulasan pengguna. Setelah tahap *preprocessing* selesai, data yang telah diproses kemudian disimpan dan siap untuk digunakan dalam tahap pelatihan model.

### 4.2.3 Pelatihan Model

Pada tahap pelatihan model, data yang telah melalui *preprocessing* dilakukan *split data* menjadi data latih dan data uji, kemudian diproses menggunakan *BertTokenizer* dari *Hugging Face Transformers* untuk mengonversi

teks menjadi format token yang bisa dipahami oleh model BERT. Setelah data tokenisasi siap, dataset kemudian diubah menjadi format yang dapat diproses oleh *PyTorch* menggunakan *DataLoader*, yang mempermudah *batching* selama pelatihan model IndoBERT (Tabel 4.6)

Tabel 4.6 Data Split

Skenario	Rasio	Data Training	Data Testing
Skenario 1	80:20	5.600	1.400
Skenario 2	90:10	6.300	700
	80:20	5.600	1.400
	70:30	4.900	2.100

#### 4.2.4 Pengujian Model *Fine-Tuning* IndoBERT

Selama proses pelatihan model IndoBERT, data yang telah melalui tahap *preprocessing* digunakan untuk melatih model menggunakan Transformers dengan fungsi *BertForSequenceClassification.from\_pretrained()* yang memuat model IndoBERT yang telah di *fine-tune* untuk klasifikasi sentimen. Pada tahap ini, model dilatih dengan menggunakan *optimizer* AdamW, yang diatur secara manual dengan parameter seperti *learning rate*, *beta1*, *beta2*, *epsilon*, dan *weight decay*. Parameter-parameter ini diatur untuk mengoptimalkan pembaruan parameter selama pelatihan. AdamW telah dilengkapi dengan *weight decay* untuk regularisasi, yang bertujuan untuk mengurangi risiko *overfitting* dengan mencegah model menjadi terlalu kompleks.

Selama pelatihan, pembaruan parameter model dilakukan melalui tiga langkah utama: pembaruan momentum pertama, momentum kedua, dan *bias correction* untuk estimasi momentum. Pembaruan ini memungkinkan model untuk memperbaiki dan menyesuaikan bobot-bobotnya seiring waktu, sehingga

menghasilkan model yang lebih akurat. Selain itu, untuk mengukur ketidaksesuaian antara prediksi model dan label sebenarnya, *cross-entropy loss* digunakan. kemudian fungsi *softmax* diterapkan pada output model untuk mengubah skor mentah (*logits*) menjadi probabilitas kelas, yang kemudian digunakan untuk menghitung loss. Hasil loss yang lebih rendah menunjukkan bahwa model berhasil memprediksi sentimen dengan lebih akurat pada data pelatihan, sehingga model semakin mendekati kinerja optimal.

Teknik *early stopping* juga diterapkan untuk mencegah *overfitting* dan memastikan model tidak terlalu terfokus pada data pelatihan. *Early stopping* menghentikan pelatihan jika tidak ada peningkatan pada *validation loss* setelah sejumlah *epoch* tertentu, sehingga model hanya akan dilatih selama periode yang diperlukan untuk mencapai hasil terbaik.

Setelah pelatihan, model dievaluasi dengan menggunakan data pengujian untuk mengukur kinerjanya melalui *confusion matrix*, dimana dilakukan perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*. Setelah evaluasi model selesai, model yang telah di *fine-tune* disimpan menggunakan fungsi *save\_pretrained()* yang memungkinkan model untuk digunakan kembali pada ulasan pengguna baru dalam aplikasi telemedicine. Langkah-langkah ini memastikan bahwa model IndoBERT yang telah di *fine-tune* mampu memberikan prediksi sentimen yang akurat dan efisien, dengan meminimalkan risiko *overfitting* dan memastikan performa yang optimal pada data baru.

### 4.3 Hasil Uji Coba

Pada tahap ini, pengujian dilakukan sesuai dengan rancangan uji coba yang telah dijelaskan sebelumnya. Pengujian ini bertujuan untuk mengevaluasi kinerja model dalam menganalisis sentimen ulasan pengguna aplikasi telemedicine dengan mempertimbangkan berbagai skenario yang berbeda. Dalam pengujian ini, dilakukan pembagian data latih dan data uji dengan beberapa rasio, yaitu 80:20, 90:10, dan 70:30, untuk melihat pengaruh pembagian data terhadap performa model. Skenario pertama mengoptimalkan *hyperparameter learning rate* dengan menggunakan teknik *Grid Search* dalam kombinasi dengan 2 *epoch* untuk melatih model. Pada skenario kedua, model dilatih dengan jumlah *epoch* yang fleksibel, diikuti dengan penerapan *early stopping* untuk mencegah *overfitting*, dengan pengaturan pembagian data latih dan data uji yang bervariasi.

Evaluasi hasil pengujian menggunakan *confusion matrix* dilakukan untuk menilai *accuracy*, *precision*, *recall*, dan *F1-score* model berdasarkan hasil prediksi sentimen pada data uji. *Confusion matrix* memberikan gambaran yang lebih jelas tentang distribusi hasil klasifikasi dan seberapa efektif model dalam mengidentifikasi kelas-kelas sentimen yang berbeda (positif, negatif, dan netral). Hasil terbaik dari berbagai skenario ini akan dianalisis untuk menentukan kombinasi parameter dan pembagian data yang menghasilkan performa terbaik pada tugas klasifikasi sentimen, serta memberikan wawasan mengenai keandalan model dalam mengklasifikasikan sentimen ulasan pengguna aplikasi telemedicine.

### 4.3.1 Pengujian Skenario 1

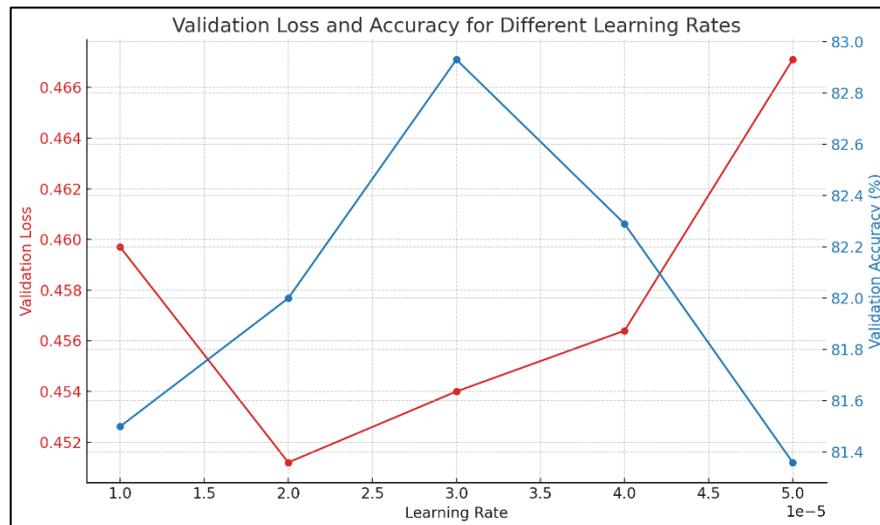
Pada pengujian dengan teknik *Grid Search* untuk optimasi *learning rate*, lima variasi *learning rate* diuji untuk menemukan nilai terbaik yang menghasilkan performa model terbaik pada data pengujian. Setiap eksperimen dilakukan dengan menggunakan 2 *epoch* pelatihan. Hasil dari eksperimen menunjukkan bahwa pada setiap *learning rate* yang diuji, nilai *train loss*, *train accuracy*, *validation loss*, dan *validation accuracy* berubah seiring dengan perubahan *learning rate*.

Berikut adalah hasil perbandingan performa model untuk masing-masing *learning rate* yang diuji pada Tabel 4.7 :

Tabel 4.7 Perbandingan *Learning rate*

No	<i>Learning rate</i>	<i>Validation loss</i>	<i>Validation accuracy</i>
1.	1e-05	0.4597	81.50%
2.	2e-05	0.4512	82.00%
3.	3e-05	0.4540	82.93%
4.	4e-05	0.4564	82.29%
5.	5e-05	0.4671	81.36%

Dari hasil tersebut, *learning rate* 3e-05 menghasilkan nilai *validation accuracy* tertinggi sebesar 82.93% dengan *validation loss* 0.4540, menjadikannya sebagai pilihan terbaik di antara variasi *learning rate* lainnya.



Gambar 4.4 Visualisasi Perbandingan *Learning rate*

Berdasarkan Tabel 4.7 dan Gambar 4.4 tersebut, *learning rate* terbaik adalah  $3e-05$  karena memiliki *validation accuracy* tertinggi sebesar 82.93%. Meskipun *learning rate*  $2e-05$  memiliki *validation loss* sedikit lebih rendah, fokus utama dalam pemilihan *learning rate* adalah *validation accuracy* sebagai indikator utama performa model.

#### 4.3.2 Pengujian Skenario 2

Pada skenario kedua, eksperimen dilakukan dengan jumlah *epoch* fleksibel dan *early stopping* untuk mencegah *overfitting* dengan ketentuan maksimal 10 *epoch*. *Early stopping* menggunakan ketentuan *patience* 5, dimana akan dilakukan penghentian pelatihan jika *validation loss* tidak membaik setelah 5 *epoch* berturut-turut. Pembagian data dilakukan dengan tiga rasio: 90:10, 80:20, dan 70:30, untuk menguji kemampuan model menggeneralisasi pada berbagai ukuran data pelatihan dan pengujian.

Hasil dari skenario pertama dengan *Grid Search* untuk *learning rate* terbaik ( $3e-05$ ) digunakan pada skenario kedua. *Learning rate* ini menghasilkan *validation accuracy* 82.93% dan *validation loss* 0.4540, yang diharapkan memberikan hasil lebih optimal pada eksperimen skenario kedua.

#### A. Split Data 90:10

Pada pengujian split data 90:10, data dibagi dengan proporsi 90% data latih dan 10% data uji, yang menghasilkan 6.300 data latih dan 700 data uji. Pada data pelatihan, terdapat distribusi sentimen yang terdiri dari 251 sentimen negatif, 253 sentimen positif, dan 196 sentimen netral. Berdasarkan percobaan yang dilakukan menggunakan *early stopping* hasil menunjukkan model terbaik berada terdapat pada *epoch* 2 (Tabel 4.8).

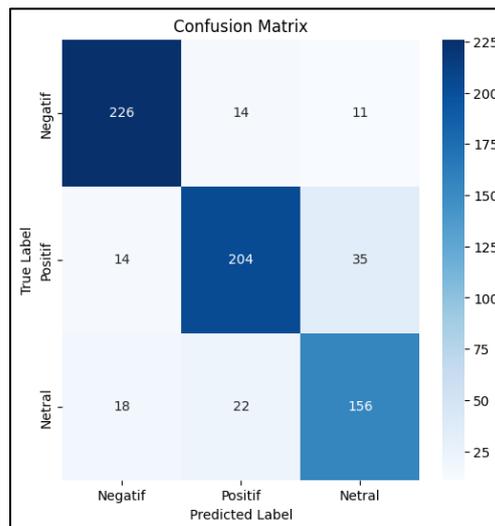
Tabel 4.8 Hasil *Early Stopping* pada split data 90:10

<i>Epoch</i>	<i>Train loss</i>	<i>Train Accuracy (%)</i>	<i>Validation loss</i>	<i>Validation accuracy (%)</i>	<b>Model Disimpan</b>
1e-05	0.5562	77.70	0.4555	82.86	Ya
2e-05	0.3878	85.59	0.4885	83.71	Ya (Terbaik)
3e-05	0.2977	88.90	0.4717	82.00	Tidak
4e-05	0.2273	91.90	0.7873	77.14	Tidak
5e-05	0.1835	93.63	0.6517	81.29	Tidak

Berdasarkan hasil dari tabel tersebut, model pelatihan berhenti pada *epoch* ke-5 karena *trigger early stopping* yang bertujuan untuk mencegah *overfitting*. Meski akurasi dan loss pada data latih terus membaik dari *epoch* pertama hingga kelima, performa model pada data validasi justru tidak konsisten. Akurasi validasi mencapai puncaknya pada *epoch* ke-2 dengan nilai 83,71%, namun kemudian menurun pada *epoch* berikutnya. Oleh karena itu, pelatihan dihentikan secara

otomatis pada *epoch* ke-5 agar model tetap optimal dan tidak mengalami penurunan performa pada data baru.

Setelah model dengan *epoch* terbaik didapatkan, maka model akan disimpan dan kemudian dilakukan evaluasi model dengan *confusion matrix*, Evaluasi dapat dilihat pada Gambar 4.5 dan Tabel 4.9.



Gambar 4.5 Confusion Matrix pada Split Data 90:10

Tabel 4.9 Hasil perhitungan *Confusion Matrix* pada Split Data 90:10

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	87.59%	90.03%	88.79%	251
Positif	85%	80.63%	82.75%	253
Netral	77.22%	79.59%	78.32%	196

$$Akurasi = \frac{226 + 204 + 156}{226 + 204 + 156 + 71 + 53} \times 100\% = 83.71\%$$

Berdasarkan hasil evaluasi pada tersebut, didapatkan akurasi sebesar 83.71%, untuk kelas negatif memiliki *precision* 87.59%, *recall* 90.03%, dan *F1-score* 88.79% dengan *support* sebanyak 251 data. Kelas positif memiliki *precision* 85%, *recall* 80.63%, dan *F1-score* 82.75 dengan *support* 253 data. Kelas netral

memiliki *precision* 77.22%, *recall* 79.59%, dan *F1-score* 78.32% dengan *support* 196 data.

#### B. Split Data 80:20

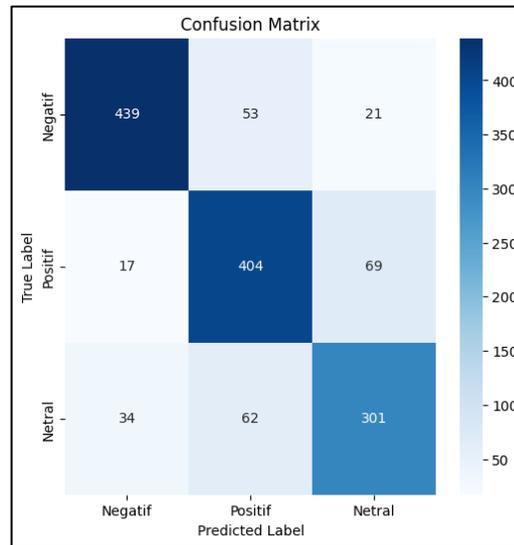
Pada pengujian split data 80:20, data dibagi dengan proporsi 80% data latih dan 20% data uji, yang menghasilkan 5.400 data latih dan 1.600 data uji. Pada data pelatihan, terdapat distribusi sentimen yang terdiri dari 513 sentimen negatif, 490 sentimen positif, dan 397 sentimen netral. Berdasarkan percobaan menggunakan mekanisme *early stopping*, hasil pelatihan menunjukkan bahwa model terbaik berada pada *epoch* 2e-05 dengan akurasi validasi tertinggi, sebagaimana terlihat pada Tabel 4.10.

Tabel 4.10 Hasil *Early Stopping* pada split data 80:20

<i>Epoch</i>	<i>Train loss</i>	<i>Train Accuracy (%)</i>	<i>Validation loss</i>	<i>Validation accuracy (%)</i>	<b>Model Disimpan</b>
1e-05	0.5645	77.83	0.5420	77.86	Tidak
2e-05	0.3992	84.98	0.4785	81.71	Ya (Terbaik)
3e-05	0.3040	89.16	0.5845	80.21	Tidak
4e-05	0.2377	91.59	0.5623	81.57	Tidak
5e-05	0.1741	93.82	0.6209	80.93	Tidak

Model pelatihan dihentikan secara otomatis pada *epoch* ke-5. Meskipun nilai *train loss* dan akurasi pada data latih membaik secara konsisten dari *epoch* pertama hingga keempat, performa model pada data validasi tidak menunjukkan peningkatan yang stabil. Akurasi validasi tertinggi tercapai pada *epoch* kedua dengan nilai 81,71%, kemudian cenderung menurun pada *epoch* berikutnya. Oleh karena itu, model terbaik disimpan pada *epoch* kedua agar performa pada data uji tetap optimal dan kemampuan generalisasi model terjaga.

Setelah model dengan *epoch* terbaik diperoleh, evaluasi dilakukan menggunakan *confusion matrix* yang dapat dilihat pada Gambar 4.6 dan Tabel 4.11.



Gambar 4.6 *Confusion Matrix* pada Split Data 70:30

Tabel 4.11 Hasil perhitungan *Confusion Matrix* pada Split Data 80:20

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	89.59%	85.57%	87.53%	513
Positif	77.84%	82.44%	80.50%	490
Netral	76.98%	75.81%	76.39%	397

$$Akurasi = \frac{439 + 404 + 301}{439 + 404 + 301 + 91 + 165} \times 100\% = 81.71\%$$

Berdasarkan hasil evaluasi tersebut, didapatkan akurasi sebesar 81.71%, untuk kelas negatif memiliki *precision* 89.59%, *recall* 85.57%, dan *F1-score* 87.53% dengan *support* sebanyak 513 data. Kelas positif memiliki *precision* 77.84%, *recall* 82.44%, dan *F1-score* 80.50% dengan *support* 490 data. Kelas netral memiliki *precision* 76.98%, *recall* 75.81%, dan *F1-score* 76.39% dengan *support* 397 data.

### C. Split Data 70:30

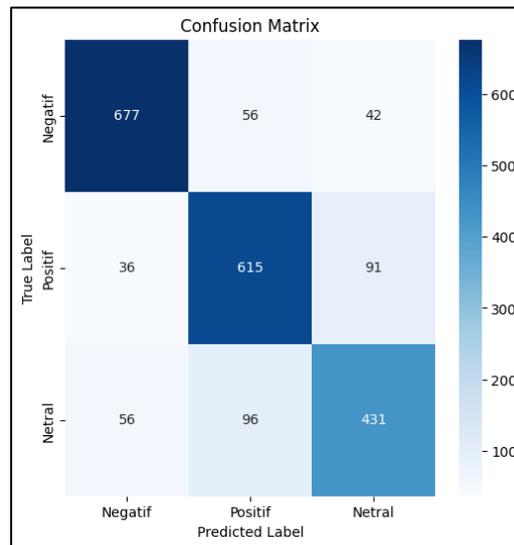
Pada pengujian split data 70:30, data dibagi dengan proporsi 70% data latih dan 30% data uji, yang menghasilkan 4.900 data latih dan 2.100 data uji. Pada data pelatihan, terdapat distribusi sentimen yang terdiri dari 775 sentimen negatif, 742 sentimen positif, dan 583 sentimen netral. Berdasarkan percobaan yang dilakukan menggunakan *early stopping*, hasil menunjukkan model terbaik terdapat pada *epoch* 3 (Tabel 4.12).

Tabel 4.12 Hasil *Early Stopping* pada split data 70:30

<i>Epoch</i>	<i>Train loss</i>	<i>Train Accuracy (%)</i>	<i>Validation loss</i>	<i>Validation accuracy (%)</i>	Model Disimpan
1e-05	0.5914	75.08	0.4987	81.33	Ya
2e-05	0.3995	85.37	0.4825	81.81	Ya
3e-05	0.2966	89.41	0.4948	82.05	Ya (Terbaik)
4e-05	0.2191	92.49	0.5962	80.81	Tidak
5e-05	0.1840	93.45	0.6619	79.62	Tidak
6e-05	0.1359	95.41	0.7597	79.48	Tidak

Berdasarkan hasil tabel tersebut, pelatihan model dihentikan pada *epoch* ke-6 karena *trigger early stopping*. Akurasi validasi mencapai nilai tertinggi pada *epoch* ke-3 sebesar 82,05%, kemudian menurun pada *epoch* berikutnya. Oleh karena itu, pelatihan dihentikan secara otomatis pada *epoch* ke-6 agar model tetap optimal dan tidak mengalami penurunan performa pada data baru.

Setelah model dengan *epoch* terbaik didapatkan, model disimpan dan kemudian dilakukan evaluasi dengan *confusion matrix*. Evaluasi dapat dilihat pada Gambar 4.7 dan Tabel 4.13.

Gambar 4.7 *Confusion Matrix* pada Split Data 70:30Tabel 4.13 Hasil perhitungan *Confusion Matrix* pada Split Data 70:30

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Negatif	88.03%	87.35%	87.68%	775
Positif	80.18%	82.88%	81.50%	742
Netral	76.41%	73.92%	75.14%	583

$$Akurasi = \frac{677 + 615 + 431}{677 + 615 + 431 + 134 + 243} \times 100\% = 82.05\%$$

Berdasarkan hasil evaluasi tersebut, didapatkan akurasi sebesar 82.05%, untuk kelas negatif memiliki *precision* 88.03%, *recall* 87.35%, dan *F1-Score* 87.68% dengan *support* sebanyak 775 data. Kelas positif memiliki *precision* 80.18%, *recall* 82.88%, dan *F1-score* 81.50% dengan *support* 742 data. Sedangkan kelas netral memiliki *precision* 76.41%, *recall* 73.92%, dan *F1-score* 75.14% dengan *support* 583 data.

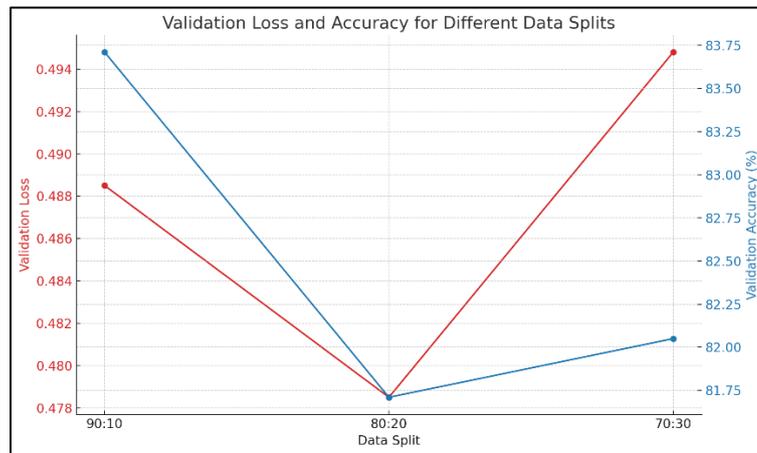
#### 4.4 Pembahasan

Berdasarkan hasil eksperimen yang dilakukan pada seluruh skenario, beberapa kesimpulan penting dapat ditarik mengenai kinerja model *Fine-tuning* IndoBERT dalam mengklasifikasikan sentimen pada data ulasan aplikasi telemedicine.

Pada Skenario 1, yang mengoptimalkan *learning rate* menggunakan teknik Grid Search, hasil eksperimen menunjukkan bahwa nilai *learning rate*  $3e-05$  memberikan performa terbaik dengan *validation accuracy* sebesar 82.93% dan *validation loss* sebesar 0.4540. Performa ini lebih unggul dibandingkan dengan variasi *learning rate* lainnya, yang menghasilkan akurasi lebih rendah. Pengujian ini menegaskan pentingnya memilih *learning rate* yang optimal untuk mencapai model dengan performa yang lebih baik.

Selanjutnya, pada Skenario 2, yang menguji jumlah *epoch* fleksibel dengan penerapan *early stopping* untuk mencegah *overfitting*, ditemukan bahwa penggunaan *learning rate* terbaik dari Skenario 1 ( $3e-05$ ) membawa model untuk memberikan hasil yang lebih optimal. Dengan menggunakan *early stopping* yang diberlakukan pada 5 *epoch* berturut-turut tanpa perbaikan pada *validation loss*, model dapat menghentikan pelatihan lebih awal dan mencegah *overfitting*.

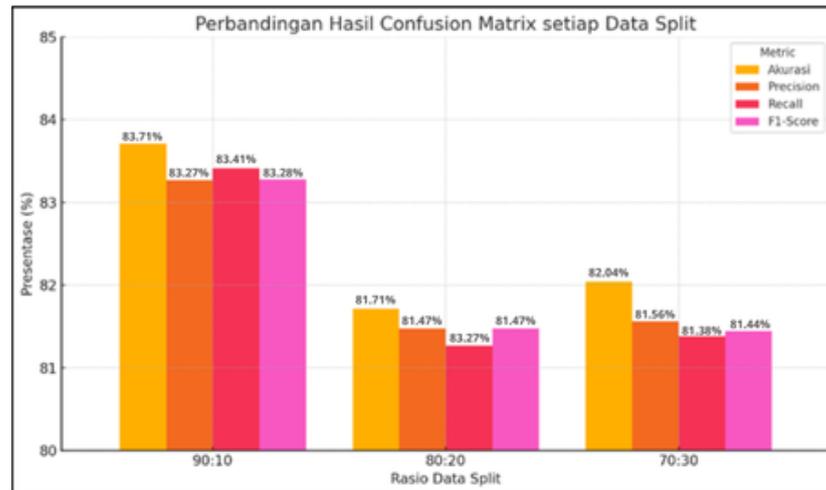
Hasil uji coba dalam perbandingan 3 rasio berdasarkan *validation loss* dan *validation accuracy* pada pembagian data latih dan data uji pada skenario 2 menggunakan *Fine-tuning* IndoBERT ditampilkan pada Gambar 4.8.



Gambar 4.8 Visualisasi berdasarkan *validation loss* dan *validation accuracy*

Berdasarkan hasil pada skenario 2, dengan membandingkan *validation loss* dan *validation accuracy*. Hasilnya menunjukkan bahwa semakin kecil rasio data uji seperti pada split data 90:10, model menghasilkan *validation accuracy* yang lebih tinggi sebesar 83.71% dan *validation loss* yang lebih rendah. Sebaliknya, pada rasio data uji yang lebih besar seperti 80:20 dan 70:30, meskipun *validation loss* juga menurun, terjadi penurunan juga pada *validation accuracy*. Temuan ini menunjukkan bahwa pembagian data dan penggunaan *early stopping* memiliki dampak signifikan terhadap kinerja model, dengan rasio 90:10 memberikan hasil terbaik dalam hal akurasi dan efisiensi pelatihan. Penelitian ini relevan dengan teori-teori dalam bidang machine learning yang menekankan pentingnya optimasi parameter dan penggunaan teknik seperti *early stopping* untuk meningkatkan kinerja model dalam skenario nyata.

Hasil uji coba dalam perbandingan 3 rasio dengan *Confusion Matrix* pada pembagian data latih dan data uji pada skenario 2 menggunakan *Fine-tuning* IndoBERT ditampilkan pada Gambar 4.9.



Gambar 4.9 Visualisasi perbandingan *Confusion Matrix* setiap Split Data

Tabel 4.14 Hasil Rata-Rata *Confusion Matrix* pada setiap rasio

Rasio	Epoch	Akurasi	Precision	Recall	F1-Score
90:10	2e-05	83.71%	83.27%	83.41%	83.28%
80:20	2e-05	81.71%	81.47%	81.27%	81.47%
70:30	3e-05	82.04%	81.56%	81.38%	81.44%

Berdasarkan hasil pada gambar 4.9 dan Tabel 4.14, pada pembagian data dengan rasio 90:10, model menunjukkan performa terbaik dengan *accuracy* sebesar 83.71%, *precision* 83.27%, *recall* 83.41%, dan *F1-score* 83.28%, yang menunjukkan keseimbangan yang baik antara kemampuan model dalam mengklasifikasikan data dengan benar. Rasio ini memberikan jumlah data latih yang cukup besar, yang memungkinkan model untuk mempelajari pola-pola dalam data dengan lebih efektif dan memberikan kinerja yang lebih stabil dan lebih tinggi dibandingkan dengan rasio lainnya. Dengan demikian, rasio 90:10 terbukti memberikan hasil terbaik dalam mengoptimalkan kinerja model, menjadikannya pilihan yang lebih ideal dalam pelatihan model *machine learning*.

Dari hasil pengujian dan evaluasi yang dilakukan, dapat disimpulkan bahwa kombinasi antara pembagian data yang tepat, optimasi *learning rate*, serta penerapan *early stopping* menghasilkan model IndoBERT yang efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi telemedicine. Pembagian data dengan rasio 90:10 memberikan performa terbaik dengan *validation accuracy* mencapai 83.71%. Optimasi *learning rate* dengan nilai  $3e-05$  terbukti memberikan hasil terbaik pada model, dengan meningkatkan *validation accuracy* dan menghindari *overfitting*. Secara keseluruhan, model yang telah di *fine-tune* menunjukkan kemampuan yang baik dalam mengidentifikasi sentimen negatif, positif, dan netral, dengan hasil evaluasi yang cukup baik pada semua kelas.

. Keberhasilan model dalam mengklasifikasikan sentimen sangat dipengaruhi oleh beberapa faktor utama, di antaranya pemilihan *learning rate* yang optimal, pembagian data yang tepat antara data latih dan data uji, serta tahap *preprocessing* teks yang dilaksanakan sebelum pelatihan. Proses *preprocessing* memiliki peran yang krusial dalam meningkatkan kualitas data yang digunakan oleh model, memungkinkan model untuk lebih fokus pada informasi yang relevan dan bermakna. Selain itu, pembagian data latih dan uji yang seimbang membantu menghindari masalah *overfitting* atau *underfitting*, yang sangat penting untuk memastikan bahwa model dapat menggeneralisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya.

Dengan menggunakan model IndoBERT yang telah di *fine-tune*, sistem ini mampu mendeteksi dan mengklasifikasikan sentimen pengguna menjadi tiga kategori utama, yaitu positif, negatif, dan netral. Pengklasifikasian ini dilakukan

dengan harapan agar dapat menghasilkan tingkat validasi akurasi yang tinggi pada ulasan yang diberikan oleh pengguna aplikasi telemedicine. Untuk mencapai tujuan tersebut, penelitian ini melakukan optimasi terhadap *learning rate* dengan menggunakan teknik Grid Search, serta menerapkan *early stopping* untuk memastikan proses pelatihan berhenti pada waktu yang tepat dan menghindari pelatihan yang berlarut-larut, yang dapat memicu *overfitting*.

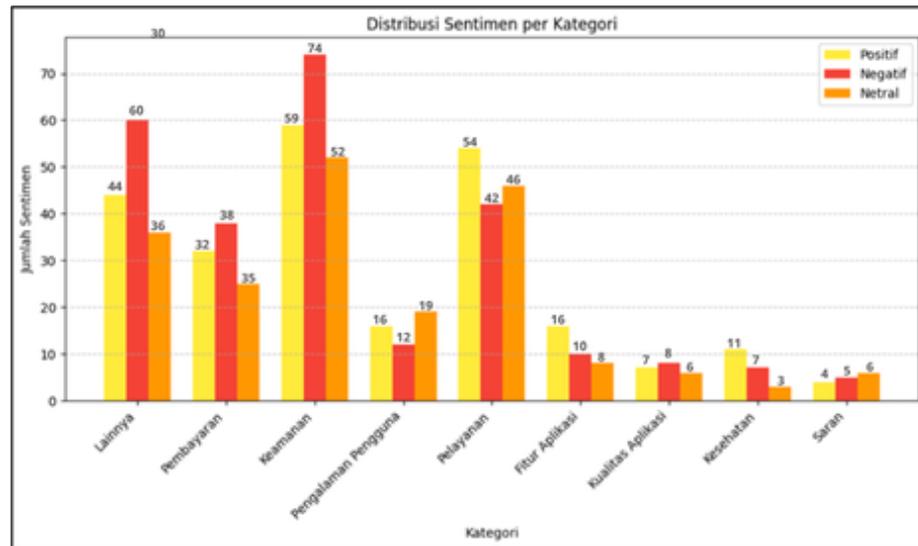
Selanjutnya, untuk mengevaluasi kinerja model yang dihasilkan, penelitian ini menggunakan *confusion matrix* yang memungkinkan pengukuran terhadap *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi komprehensif ini memberikan gambaran yang lebih jelas mengenai efektivitas model dalam mengidentifikasi dan mengklasifikasikan sentimen yang ada dalam ulasan pengguna. Hasil dari evaluasi menunjukkan bahwa model IndoBERT yang telah di-fine-tune menunjukkan kinerja yang cukup baik dalam mengidentifikasi sentimen negatif, dengan nilai *precision* dan *recall* yang tinggi pada kategori ini. Meskipun demikian, terdapat beberapa ruang untuk peningkatan pada pengklasifikasian sentimen positif dan netral.

Secara keseluruhan, penelitian ini mengembangkan sebuah model klasifikasi sentimen yang robust dan akurat untuk ulasan pengguna aplikasi telemedicine. Model *fine-tuning* IndoBERT berhasil menunjukkan performa yang lebih baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi telemedicine. Teknik *early stopping* terbukti efektif dalam mencegah *overfitting*, sehingga model dapat mencapai akurasi optimal tanpa kehilangan kemampuan generalisasi. Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang

berguna bagi penyedia layanan telemedicine untuk meningkatkan layanan mereka berdasarkan umpan balik yang lebih objektif dan terperinci dari pengguna. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam hal pengembangan teknologi analisis sentimen, tetapi juga memberikan nilai praktis untuk pengembangan aplikasi dan peningkatan pengalaman pengguna secara keseluruhan.

#### **4.5 Visualisasi Kategori pada Ulasan**

Pada bagian ini dilakukan pengkategorian ulasan yang diambil dari hasil split data terbaik 90:10, dengan tujuan untuk mengelompokkan ulasan berdasarkan kategori-kategori spesifik, seperti "Keamanan", "Pelayanan", "Fitur Aplikasi", dan lainnya, yang ditentukan berdasarkan kata kunci yang relevan. Proses ini bertujuan untuk memberikan pemahaman yang lebih terstruktur mengenai sentimen yang terkandung dalam ulasan pengguna. Setiap ulasan dianalisis untuk menentukan kategori yang paling sesuai berdasarkan kemunculan kata kunci yang relevan, kemudian dikelompokkan dalam tiga sentimen utama: positif, negatif, dan netral. Berikut visualisasi hasil yang didapat dari model, terdapat pada Gambar 4.10 dan Tabel 4.15.



Gambar 4.10 Visualisasi hasil kategori pada

Tabel 4.15 Hasil pengkategorian ulasan

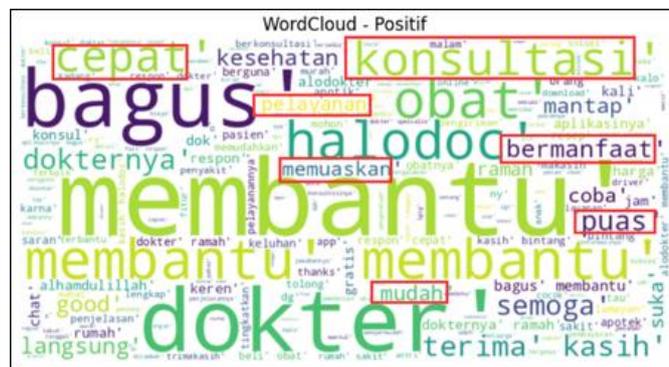
No.	Kategori	Positif	Negatif	Netral
1.	Keamanan	59	74	52
2.	Pelayanan	54	42	46
3.	Pembayaran	32	38	25
4.	Pengalaman Pengguna	16	12	19
5.	Fitur Aplikasi	16	10	8
6.	Kesehatan	11	7	3
7.	Kualitas Aplikasi	7	8	6
8.	Saran	4	5	6
9.	Lainnya	44	60	36
<b>Total</b>		<b>243</b>	<b>256</b>	<b>201</b>
<b>Total Keseluruhan</b>		<b>700</b>		

Berdasarkan hasil tersebut menunjukkan bahwa kategori keamanan dan pelayanan mendominasi dengan sentimen negatif, masing-masing dengan 74 dan 42 ulasan negatif. Kategori Keamanan juga memiliki sentimen positif yang signifikan, yakni 59, sementara kategori pelayanan memiliki keseimbangan yang lebih seimbang antara positif dan negatif. Kategori fitur aplikasi cenderung lebih positif, dengan 16 ulasan positif dibandingkan 10 ulasan negatif. Kategori kesehatan menunjukkan sedikit sentimen negatif dan lebih banyak ulasan positif.

Secara keseluruhan, sentimen negatif banyak muncul pada kategori yang berkaitan dengan masalah keamanan dan pelayanan, sementara fitur aplikasi dan kesehatan lebih banyak diterima positif.

#### 4.6 Visualisasi Kata pada Ulasan

Visualisasi ini dilakukan menggunakan *WordCloud*. *WordCloud* adalah sebuah teknik visualisasi data yang menampilkan kata-kata yang sering muncul dalam suatu kumpulan teks. Dalam visualisasi ini, kata yang lebih sering muncul akan memiliki ukuran yang lebih besar. *WordCloud* sangat berguna dalam analisis sentimen karena memberikan gambaran yang jelas tentang kata-kata yang dominan dalam suatu teks. Dalam konteks sentimen *WordCloud* dapat membantu untuk menyoroti kata-kata yang paling sering digunakan dalam ulasan pelanggan positif, negatif, dan netral.



Gambar 4.11 *WordCloud* Sentimen Positif

Berdasarkan Gambar 4.11 visualisasi pada sentimen positif, kata yang sering muncul dalam sentimen positif adalah kata-kata yang mencerminkan kepuasan dan apresiasi terhadap layanan atau produk yang diterima. Terdapat 5 kata yang paling dominan pada visualisasi *WordCloud* sentimen positif (Gambar 4.12):

Top 5 Kata pada Sentimen Positif: Top 5 Kata: membantu: 879 dokter: 401 bagus: 302 obat: 216 cepat: 183
---

Gambar 4.12 Top 5 Kata pada Sentimen Positif

“membantu” yang muncul sebanyak 879 kali, . Hal ini menunjukkan bahwa banyak pelanggan merasa sangat terbantu oleh layanan yang diberikan. Selain itu, Kata-kata seperti “membantu,” “bagus”, “bermanfaat”, dan “puas” menunjukkan bahwa pelanggan merasa dihargai dan merasa bahwa layanan yang mereka terima sangat bermanfaat hal ini menunjukkan bahwa banyak responden memberikan penilaian positif terhadap kualitas layanan atau produk. Kata “konsultasi” dan “pelayanan” juga muncul dalam visualisasi ini, yang mengindikasikan pentingnya interaksi antara pelanggan dan penyedia layanan. Kata “mudah” dan “cepat” juga muncul dalam visualisasi ini, hal ini menunjukkan bahwa proses yang dijalani oleh pelanggan, baik itu dalam mendapatkan layanan atau menggunakan produk, berlangsung tanpa hambatan atau kerumitan. Hal ini mengarah pada pengalaman pengguna yang nyaman dan efisien.

Secara keseluruhan, berdasarkan visualisasi ini, dapat disimpulkan bahwa sentimen positif yang dominan terkait dengan kualitas layanan yang membantu, konsultasi yang berguna, dan kepuasan pelanggan. Visualisasi ini memperlihatkan bahwa kualitas pelayanan yang baik dan komunikasi yang efektif adalah faktor utama yang membangun sentimen positif di kalangan pelanggan.





Kata “coba” juga muncul sebanyak 491 kali, yang mengindikasikan adanya keinginan dari pelanggan untuk mencoba produk atau layanan tersebut, namun tanpa penekanan yang kuat pada hasil akhir, baik positif maupun negatif. Kata “semoga” yang juga muncul menambah kesan bahwa banyak ulasan yang mengandung harapan atau ekspektasi terhadap layanan atau produk yang belum sepenuhnya tercapai. Ini menunjukkan bahwa meskipun tidak ada keluhan yang jelas, pelanggan berharap agar layanan atau produk dapat berkembang atau diperbaiki di masa depan. Kata “bintang” muncul dengan frekuensi cukup tinggi, yang menunjukkan bahwa banyak pengguna memberikan ulasan atau penilaian terhadap layanan atau produk dengan memberi rating atau tanda bintang. Hal ini mengindikasikan bahwa responden lebih fokus pada memberi masukan atau rekomendasi tanpa menyatakan ketidakpuasan atau kepuasan secara ekstrem.

Secara keseluruhan, visualisasi ini menunjukkan bahwa sentimen netral lebih berfokus pada harapan, penilaian umum, dan keinginan untuk mencoba tanpa adanya klaim kepuasan atau ketidakpuasan yang kuat. Hal ini mencerminkan bahwa pelanggan tidak sepenuhnya puas atau kecewa, tetapi lebih memberi umpan balik yang konstruktif atau netral terhadap layanan atau produk yang mereka coba.

#### **4.7 Integrasi Islam**

Penelitian ini mengkaji persepsi masyarakat terhadap aplikasi telemedicine melalui pendekatan analisis sentimen berbasis metode *fine-tuning*, dengan mengintegrasikan nilai-nilai Islam sebagai landasan moral dan spiritual. Dalam perspektif Islam, kesehatan merupakan aspek penting yang mencakup dimensi fisik,

mental, dan spiritual, sehingga pengobatan tidak hanya dipahami sebagai proses medis, tetapi juga sebagai bentuk pengabdian dan kepercayaan kepada Allah SWT.

Prinsip ini diperkuat oleh sabda Rasulullah SAW dalam hadis yang diriwayatkan oleh Imam Bukhari (No. 5354):

مَا أَنْزَلَ اللَّهُ دَاءً إِلَّا أَنْزَلَ لَهُ شِفَاءً

*“Tidaklah Allah menurunkan suatu penyakit kecuali Dia juga menurunkan penawarnya.” (HR. Bukhari no.5354)*

Dalam kitab Shahih Bukhari Muslim, Hadis ini menjadi landasan teologis bahwa setiap penyakit memiliki solusi yang dapat ditemukan melalui penelitian dan inovasi, termasuk dalam pengembangan layanan kesehatan berbasis teknologi. Maka, pemanfaatan telemedicine dan analisis sentimen dalam penelitian ini bukan hanya bersifat teknis, tetapi juga berkontribusi pada nilai kemanusiaan dan tanggung jawab etis sesuai ajaran Islam.

Dalam penelitian ini, pengintegrasian nilai-nilai Islam memberikan dimensi moral dan spiritual terhadap penggunaan teknologi dalam bidang kesehatan. Dalam perspektif Islam, teknologi tidak hanya dipandang sebagai alat untuk mencapai tujuan duniawi, tetapi juga sebagai sarana yang menghubungkan manusia dengan Allah SWT dalam upaya mencapai kesembuhan dan kesejahteraan, yang dalam terminologi fiqih disebut sebagai *Muamalah Ma'a Allah*, yaitu interaksi manusia dengan Sang Pencipta dalam rangka mencapai kesembuhan dan kesejahteraan. Konsep ini sejalan dengan ajaran Islam yang menekankan bahwa segala bentuk ikhtiar, baik yang bersifat fisik maupun digital, harus dilakukan dengan penuh kesadaran bahwa hasil akhirnya tetap berada dalam takdir dan kehendak Allah

SWT. Melalui analisis sentimen terhadap aplikasi telemedicine, kita tidak hanya mengukur persepsi pengguna terhadap kualitas layanan yang diberikan, tetapi juga mencoba menilai sejauh mana masyarakat memahami dan mengapresiasi hubungan antara ikhtiar manusia dan ketentuan Tuhan dalam proses penyembuhan.

Lebih lanjut, pendekatan ini menekankan pentingnya hubungan sosial atau *Muamalah Ma'a Annas*, yakni hubungan antar sesama manusia, di mana aplikasi telemedicine tidak hanya memberikan manfaat individu dalam memperoleh perawatan medis, tetapi juga berkontribusi pada tanggung jawab sosial untuk meningkatkan kualitas hidup dan kesehatan umat. Hal ini sejalan dengan firman Allah dalam Al-Qur'an Surah Al-Ma'idah [5]: 2:

وَتَعَاوَنُوا عَلَى الْبِرِّ وَالتَّقْوَىٰ وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ ﴿٢﴾

*“Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan” (Q.S Al-Ma'idah)*

Menurut tafsir Ibnu Katsir yang dikutip pada website Tanwir yang ditulis oleh Nikma (2023), ayat ini menegaskan kewajiban umat Islam untuk saling membantu dalam hal-hal yang membawa kebaikan dan ketakwaan bukan kemungkarannya. Tolong-menolong yang dimaksud tidak sebatas bantuan fisik, tetapi juga dukungan moral, sosial, dan spiritual yang mempererat ukhuwah dan solidaritas antarindividu dalam masyarakat. Dengan demikian, penggunaan teknologi kesehatan seperti telemedicine menjadi bagian dari implementasi prinsip *Muamalah Ma'a An-nas* yang mendukung upaya bersama dalam menjaga dan meningkatkan kesehatan umat secara menyeluruh.

Islam mengajarkan bahwa menjaga kesehatan adalah bagian dari ibadah, dan setiap upaya untuk mengakses perawatan medis, baik melalui cara konvensional maupun digital, merupakan bagian dari usaha untuk menjaga amanah tubuh yang diberikan oleh Allah. Oleh karena itu, penelitian ini tidak hanya mengevaluasi aspek teknis dan ilmiah dari aplikasi telemedicine, tetapi juga menilai sejauh mana teknologi ini dapat mendukung upaya umat manusia dalam mencapai kesehatan yang lebih baik sesuai dengan ajaran Islam. Pendekatan yang menggabungkan aspek spiritual, moral, dan teknis ini memberikan perspektif yang lebih komprehensif tentang bagaimana teknologi dapat digunakan secara etis dan bertanggung jawab dalam meningkatkan kualitas hidup masyarakat, dengan tetap menjaga hubungan yang baik dengan Allah dan sesama manusia.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil analisis dan eksperimen yang telah dilakukan, penelitian ini berhasil menunjukkan bahwa pendekatan fine-tuning pada model IndoBERT efektif dalam mengklasifikasikan sentimen pengguna terhadap aplikasi telemedicine Halodoc dan Alodokter. Dengan memanfaatkan 7.000 ulasan pengguna berbahasa Indonesia dan menerapkan skenario optimasi hyperparameter, penelitian ini menemukan bahwa penggunaan learning rate sebesar  $3e-05$  menghasilkan akurasi validasi tertinggi sebesar 82,93% serta nilai loss sebesar 0.4885. Selain itu, penerapan teknik early stopping dengan patience sebanyak 5 epoch mampu mencegah overfitting dan meningkatkan efisiensi proses pelatihan model.

Ekperimen juga menunjukkan bahwa rasio pembagian data latih dan uji sebesar 90:10 memberikan performa terbaik dengan akurasi validasi mencapai 83,71%, precision 83,27%, recall 83,41%, dan F1-score sebesar 83,28%. Temuan dalam proses pelatihan model pembelajaran mesin. Secara keseluruhan, kombinasi dari pemilihan learning rate yang optimal, penerapan early stopping, dan pembagian data yang tepat terbukti berpengaruh signifikan terhadap kinerja model IndoBERT dalam menganalisis opini pengguna.

## 5.2 Saran

Peneliti menyadari bahwa hasil dari penelitian ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Agar sistem yang dikembangkan dapat berfungsi secara lebih optimal dan memberikan hasil yang lebih komprehensif, diperlukan beberapa langkah pengembangan lanjutan yang dapat meningkatkan kualitas dan performa analisis. Adapun saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Mengambil data dari berbagai platform aplikasi atau media sosial untuk memperkaya konteks dan keragaman opini pengguna.
2. Eksplorasi *hyperparameter* model dapat diperdalam, tidak terbatas pada *learning rate*, tetapi juga mencakup *batch size*, arsitektur model, dan strategi *regularization* tambahan untuk mengidentifikasi konfigurasi yang paling optimal.
3. Membandingkan performa IndoBERT dengan model lain seperti RoBERTa atau DistilBERT.
4. Mengintegrasikan anotasi tambahan seperti emosi atau topik untuk memperluas dimensi analisis.

## DAFTAR PUSTAKA

- Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT | Jurnal JTIC (Jurnal Teknologi Informasi dan Komunikasi)*. (t.t.). Diambil 21 April 2025, dari <https://journal.lembagakita.org/index.php/jtik/article/view/3340>
- Alfarobby, A. N. (2023). *Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online Pada Ulasan Google Playstore Menggunakan Indobert Dan Topic Modeling (Studi Kasus: Gojek Dan Grab)*. Universitas Telkom, S1 Manajemen (Manajemen Bisnis Telekomunikasi & Informatika).
- Anam, M. K., Defit, S., Haviluddin, H., Efrizoni, L., & Firdaus, M. B. (2024). *Early stopping on CNN-LSTM Development to Improve Classification Performance*. *Journal of Applied Data Sciences*, 5(3), 1175–1188.
- Bakhtiar, H. S. (2022). Dikotomi Eksistensi Telemedicine Bagi Masyarakat Terpendek: Perspektif Teori Kemanfaatan. *Jurnal Paradigma: Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, 3(2), Article 2. <https://doi.org/10.22146/jpmmmpi.v3i2.79461>
- Berobat dalam Pandangan Islam*. (t.t.). NU Online. Diambil 15 Februari 2025, dari <https://nu.or.id/tasawuf-akhlak/berobat-dalam-pandangan-islam-xQgUJ>
- Cikania, R. N. (2021). *Analisis Sentimen Review Pengguna Layanan Telemedicine Halodoc Menggunakan Algoritma Naive Bayes Classifier dan Support Vector Machine* [Undergraduate, Institut Teknologi Sepuluh Nopember]. [https://repository.its.ac.id/87562/?utm\\_source=chatgpt.com](https://repository.its.ac.id/87562/?utm_source=chatgpt.com)
- [databoks.katadata.co.id](https://databoks.katadata.co.id). (t.t.). *Mayoritas Warga RI Belum Pernah Pakai Layanan Telemedis, Apa Alasannya?* | *Databoks*. Diambil 21 April 2025, dari <https://databoks.katadata.co.id/layanan-konsumen-kesehatan/statistik/0b14214ba52bbdd/mayoritas-warga-ri-belum-pernah-pakai-layanan-telemedis-apa-alasannya>
- Fatharani, F., Kania, K. P., Hutahaean, J., & Wulan, S. R. (2022). Deteksi Intensi Chatbot Berbahasa Indonesia dengan Menggunakan Metode Capsule Network. *Journal of Information System Research (JOSH)*, 3(4), Article 4. <https://doi.org/10.47065/josh.v3i4.1821>
- Geni, L., Yulianti, E., & Sensuse, D. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using BERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9, 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>

- Hara, T. J. (1976). Structure-activity relationships of amino acids in fish olfaction. *Comparative Biochemistry and Physiology. A, Comparative Physiology*, 54(1), 31–36. [https://doi.org/10.1016/s0300-9629\(76\)80067-9](https://doi.org/10.1016/s0300-9629(76)80067-9)
- Hasan, A., Ramadhan, Y., & Minarto, M. (2023). Sentiment Analysis of Telemedicine Applications on Twitter Using Lexicon-Based and Naive Bayes Classifier Methods. *Jurnal Riset Informatika*, 5, 481–490. <https://doi.org/10.34288/jri.v5i4.244>
- Imaduddin, H., A'la, F. Y., & Nugroho, Y. S. (2023). Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(8), Article 8. <https://doi.org/10.14569/IJACSA.2023.0140813>
- Indonesia, B. P. S. (t.t.). *Statistik Kesehatan 2022*. Diambil 15 Februari 2025, dari <https://www.bps.go.id/id/publication/2023/08/31/923a16f1d75232565f1e0446/statistik-kesehatan-2022.html>
- Jamin Keamanan Masyarakat Gunakan Telemedisin, Kemenkes Minta Penyedia Layanan Daftar Regulatory Sandbox. (2023, April 13). *Sehat Negeriku*. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230413/1742765/jamin-keamanan-masyarakat-gunakan-telemedisin-kemenkes-minta-penyedia-layanan-daftar-regulatory-sandbox/>
- Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, 14(3), Article 3. <https://doi.org/10.33096/ilkom.v14i3.1505.348-354>
- Komparasi Brand*. (t.t.). Diambil 15 Februari 2025, dari [https://www.topbrand-award.com/komparasi\\_brand/bandingkan?id\\_award=1&id\\_kategori=7&id\\_subkategori=854&tahun\\_awal=2023&tahun\\_akhir=2024&brand1=Alodokter&brand2=Halodoc&brand3=KlikDokter](https://www.topbrand-award.com/komparasi_brand/bandingkan?id_award=1&id_kategori=7&id_subkategori=854&tahun_awal=2023&tahun_akhir=2024&brand1=Alodokter&brand2=Halodoc&brand3=KlikDokter)
- Laurensz, B., & Sedyono, E. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 10, 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Lestari, S., & Gozali, D. (2021). Narrative Review: Telemedicine Dan Implementasinya Dalam Membantu Perawatan Pasien Covid-19. *Farmaka*, 19(3), 63–72. <https://doi.org/10.24198/farmaka.v19i3.34918>
- Mahajaya, N. S., Ayu, P. D. W., & Huizen, R. R. (2024). Pengaruh Optimizer Adam, AdamW, SGD, dan LAMB terhadap Model Vision Transformer pada Klasifikasi

Penyakit Paru-paru. *Seminar Hasil Penelitian Informatika Dan Komputer (SPINTER) | Institut Teknologi Dan Bisnis STIKOM Bali*, 818–823.

Mahardhika, A. A., Saptono, R., & Anggrainingsih, R. (2015). Sistem Klasifikasi Feedback Pelanggan Dan Rekomendasi Solusi Atas Keluhan Di Upt Puskom Uns Dengan Algoritma Naïve Bayes Classifier Dan Cosine Similarity. . . *Juni*, 4(1).

Mahira, S. A., Sukoco, I., Barkah, C. S., & Novel, N. J. A. (2023). Teknologi Artificial Intelligence Dalam Analisis Sentimen: Studi Pada Perusahaan Kata.Ai. *Responsive: Jurnal Pemikiran Dan Penelitian Administrasi, Sosial, Humaniora Dan Kebijakan Publik*, 6(2), 139–148. <https://doi.org/10.24198/responsive.v6i2.48064>

Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), Article 1. <https://doi.org/10.22146/jisph.44455>

Masruroh, F., Surarso, B., & Warsito, B. (2023). Perbandingan Kinerja Inception-Resnetv2, Xception, Inception-v3, dan Resnet50 pada Gambar Bentuk Wajah. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(1), 11–20. <https://doi.org/10.25126/jtiik.20231014941>

Mutmainah, S., Fudholi, D., & Hidayat, S. (2023). Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine Pada Google Play Menggunakan BiLSTM dan LDA. *Jurnal Media Informatika Budidarma*, 7, 312. <https://doi.org/10.30865/mib.v7i1.5486>

Nikma, S. (2023, Juni 4). Anjuran Tolong-menolong: Tafsir Surah Al-Ma'idah Ayat 2. *Tanwir.ID*. <https://tanwir.id/anjuran-tolong-menolong-tafsir-surah-al-maidah-ayat-2/>

Pangestu, R. A., Rahmat, B., & Anggraeny, F. T. (2020). *Implementasi Algoritma Cnn Untuk Klasifikasi Citra Lahan Dan Perhitungan Luas*. 1(1).

Parwanto, D. (t.t.). *Implementasi Aplikasi Telemedicine dalam Analisis Sentimen dan Pemodelan Topik*. Rri.Co.Id - Portal Berita Terpercaya. Diambil 15 Februari 2025, dari <https://www.rri.co.id/iptek/305664/implementasi-aplikasi-telemedicine-dalam-analisis-sentimen-dan-pemodelan-topik>

*Perbandingan Kinerja Inception- Resnetv2, Xception, Inception-v3, dan Resnet50 pada Gambar Bentuk Wajah | Jurnal Teknologi Informasi dan Ilmu Komputer*. (t.t.). Diambil 19 Maret 2025, dari <https://jtiik.ub.ac.id/index.php/jtiik/article/view/4941>

Putra, H. '. (2021, April 12). *Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT*. <https://www.semanticscholar.org/paper/Deteksi-Penggunaan-Kalimat-Abusive-Pada-Teks-Bahasa-Putra/f87e445000fa8625e1bda6c904ca2b666dee0fee>

- Putri, A. M., Nofa, W. K., & Hapsari, D. A. P. (2025). Penerapan Metode Bert Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Segari Di Google Play Store. *Jurnal Ilmiah Teknik*, 4(1), Article 1. <https://doi.org/10.56127/juit.v4i1.1902>
- Ramadhan, F., & Hernadi, J. (2025). Evaluasi Optimizer Adam dan RMSProp pada Arsitektur VGG-19 Klasifikasi Ekspresi Wajah Manusia. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(2), Article 2. <https://doi.org/10.29100/jipi.v10i2.6197>
- Rufaida, A. S. R. (2023). *Analisis Sentimen Opini Masyarakat terhadap Platform Telemedicine menggunakan Metode Kamus Sentimen Gabungan dan Support Vector Machine* [Universitas Gadjah Mada]. <https://etd.repository.ugm.ac.id/penelitian/detail/228138>
- Singgalen, Y. A. (2025). Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data. *Journal of Information System Research (JOSH)*, 6(2), Article 2. <https://doi.org/10.47065/josh.v6i2.6505>
- Surat Asy-Syu'ara Ayat 80 Arab, Latin, Terjemah dan Tafsir | Baca di *TafsirWeb*. (t.t.). Diambil 19 Mei 2025, dari <https://tafsirweb.com/6490-surat-asy-syuara-ayat-80.html>
- Sutoyo, E., & Permana, M. C. (2024). *Telemedicine-Sentiment*. 1. <https://doi.org/10.17632/j8yjvsyr9p.1>
- Tarwoto, Nugroho, R., Azka, N., & Graha, W. S. R. (2025). Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 9(2), Article 2. <https://doi.org/10.35870/jtik.v9i2.3340>
- Wabang, K., Nurhayati, O., & Farikhin. (2022). Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6, 872–876. <https://doi.org/10.29207/resti.v6i5.4498>

